

UNIVERSIDADE FEDERAL DO PARANÁ
MESTRADO EM BIOLOGIA CELULAR E MOLECULAR

EDUARDO LEMONS FRANCISCO

ANÁLISE DO PADRÃO DE DECAIMENTO
DO TRANSCRIPTOMA DE *TRYPANOSOMA CRUZI*

CURITIBA

2011

EDUARDO LEMONS FRANCISCO

ANÁLISE DO PADRÃO DE DECAIMENTO
DO TRANSCRIPTOMA DE *TRYPANOSOMA CRUZI*

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Biologia Molecular, no Curso de Pós-Graduação em Biologia Celular e Molecular, Departamento de Biologia Celular, Setor de Ciências Biológicas, Universidade Federal do Paraná.

Orientação: Dr. Christian Macagnan Probst.

CURITIBA

2011

Universidade Federal do Paraná
Sistema de Bibliotecas

Francisco, Eduardo Lemons

Análise do padrão de decaimento do transcriptoma de *Trypanosoma cruzi*. / Eduardo Lemons Francisco. – Curitiba, 2011.

93 p.: il. ; 30cm.

Orientador: Christian Macagnan Probst

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

1. Tripanossomo I. Título II. Probst, Christian Macagnan III. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Biologia Celular e Molecular.

CDD (20. ed.) 616.9363

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Departamento de Biologia Celular e Departamento de Fisiologia

Setor de Ciências Biológicas

Universidade Federal do Paraná

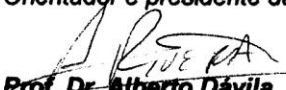
Instituto Carlos Chagas (ICC/FIOCRUZ)

PARECER

A banca examinadora, instituída pelo colegiado do Programa de Pós-Graduação em Biologia Celular e Molecular, do Setor de Ciências Biológicas, da Universidade Federal do Paraná, composta por:


Prof. Dr. Christian Macagnan Probst

Orientador e presidente da banca do Instituto Carlos Chagas - ICC


Prof. Dr. Alberto Dávila
FIOCRUZ - RJ


Prof. Dr. Fabrício Marchini
Instituto Carlos Chagas - ICC

E tendo como suplente,

Profa. Dra. Daniela P Pavoni


Instituto Carlos Chagas - ICC

Após argüir o mestrando **Eduardo Lemons Francisco**, em relação ao seu trabalho intitulado: "Análise do padrão de decaimento do transcriptoma de *trypanosoma cruzi*," são de parecer favorável à aprovação do acadêmico, habilitando-o ao título de MESTRE em Biologia Celular e Molecular.

A obtenção do título está condicionada à implementação das correções sugeridas pelos membros da banca examinadora e ao cumprimento integral das exigências estabelecidas no Regimento interno deste Programa de Pós-Graduação.

Curitiba, 28 de fevereiro de 2011




Prof. Dr. Edvaldo da Silva Trindade
Coordenador do Programa de Pós-Graduação
em Biologia Celular e Molecular - UFPR
Matr. 185795

Aos meus pais, irmão e demais familiares.
A todas as pessoas que certa vez cruzaram meu caminho.

AGRADECIMENTOS

Devo agradecimentos ao meu orientador pelas valiosas lições sobre como a ciência deve ser feita, como os questionamentos devem nos guiar, como uma paixão inabalável pelo conhecer deve ser mantida a todo custo e, principalmente, pelo exemplo de profissional dedicado e por acreditar fielmente em meu potencial.

Agradeço também aos meus colegas de trabalho pela companhia, amizade e ensinamentos nessa caminhada que abriu novas trilhas. Sinto-me bastante grato para com a Rita, cujo auxílio foi crucial durante as etapas laboratoriais desse trabalho, devido ao seu conhecimento e à sua grande prática em cultivo de células e biologia molecular.

Meu muito obrigado aos amigos e familiares que sempre me apoiaram. E, na etapa final, um agradecimento especial a Larissa, minha grande companheira, além de colega, cujo apoio incansável e paciência foram essenciais para que esse trabalho fosse concretizado e para que a fé creditada a mim pelo meu orientador e por todos fosse passível de justificativa.

RESUMO

O protozoário *Trypanosoma cruzi* controla sua expressão gênica principalmente por mecanismos pós-transcricionais como, por exemplo, a estabilidade de mRNAs. Em trabalho anterior, foram realizados experimentos de microarranjos, para avaliar o decaimento de mRNA de epimastigotas de *T. cruzi* Dm28c, tratados por 15, 30, 45, 60, 90, 120, 150, 180 e 240 minutos com 10 µg/mL e 50 µg/mL de actinomicina D (pontos em replicata, totalizando 40 amostras). Comparando-se os pontos temporais extremos, observou-se de 2 a 3 mil genes diferencialmente expressos para estabilidade de mRNA. Devido às limitações inerentes à técnica de microarranjos, e à possibilidade de se utilizar sequenciamento de DNA como abordagem, a qual possui diversas vantagens em relação aos microarranjos, foi idealizado o presente trabalho, que foi replanejado de acordo com as informações colhidas no trabalho citado acima. Foram analisados pontos temporais de decaimento menores (0, 5, 10, 15, 20 e 25 minutos) com a técnica de sequenciamento *RNA-Seq WT*. As vantagens em relação aos microarranjos são: medida absoluta (quantitativa), maiores alcance dinâmico, resolução e cobertura genômica, sem hibridização cruzada e com pouco ruído de fundo. A dose efetiva do inibidor actinomicina foi fixada em 10 µg / mL / 10⁷ células e foi acrescido sinefungina a 1 µg / mL / 10⁷ células para as análises de decaimento dos mRNAs de *T. cruzi*. Em conjunto, nossos dados mostram que uma pequena parte da regulação da expressão gênica desse parasito durante o seu ciclo de vida pode ser explicada pela estabilidade de suas moléculas de mRNA. Já que, foi possível relacionar a velocidade de decaimento dos transcritos maduros com grupos funcionais envolvidos com montagem e manutenção do citoesqueleto, de ribossomos e de outros tipos de complexos ribonucleoproteicos.

Palavras-chave: Transcriptoma; *RNA-Seq WT*; *Trypanosoma cruzi*; epimastigota.

ABSTRACT

The protozoan *Trypanosoma cruzi* controls its gene expression mainly by post-transcriptional mechanisms such as mRNA stability. In a previous work, microarray experiments were prepared to evaluate the mRNA decay in *T. cruzi* Dm28c epimastigotes, treated for 15, 30, 45, 60, 90, 120, 150, 180 and 240 minutes with 10 µg/mL and 50 µg/mL of actinomycin D. Each experimental unity was made in replicate, totalizing 40 samples. Comparing the extreme time points, 2 to 3 thousand differentially expressed genes were identified, concerning molecule stability. This data are a rich repository for mRNA decay analysis, however, the need to hybridize nucleic acid targets and probes, as well as the relative characteristic of the measure and the dependency of signal/background noise ratio produce less conclusive results. In this way, the present project covers earlier (0, 5, 10, 15, 20 and 25 minutes) mRNA decay time points and utilizes the *RNA-Seq WT* sequencing technique. Among the advantages of RNA-Seq, the measure is quantitative (absolute instead of relative) with better resolution and probe-independent, which confers wider genomic coverage, no cross-hybridization and smaller background noise. Taken together our experiments show that the differentially expressed genes during *T. cruzi*'s life cycle in epimastigote forms can be partially explained by mRNA decay rates. Moreover, it is possible to correlate mature transcript decay rates with some gene functions resembling cytoskeleton and organelle organization. Mainly genes related with macromolecular complexes such as ribosomes and ribonucleoproteins.

Keywords: Transcriptome; RNA-Seq WT; *Trypanosoma cruzi*; epimastigote.

LISTA DE FIGURAS

Figura 2.1. Ciclos silvestre (A) e doméstico (B) dos triatomíneos e do protozoário <i>T. cruzi</i> .	19
Figura 2.2. O ciclo de vida de <i>T. cruzi</i> (simplificado).	20
Figura 2.3. Fórmula química planar da molécula actinomicina D.	30
Figura 2.4. Modelo tridimensional da interação de actinomicina D com a molécula de DNA.	30
Figura 4.1. Organograma dos tipos de mapeamentos testados.	38
Figura 5.1. Perfil eletroforético comum às amostras de RNA total.	44
Figura 5.2. Histograma dos valores de " R^2 " para as amostras da cepa Y.	46
Figura 5.3. Histograma dos valores de " R^2 " para as amostras da cepa Dm28c.	46
Figura 5.4. Densidade de sequências cuja porcentagem de bases apresenta qualidade maior que 15.	49
Figura 5.5. Número de genes de acordo com suas médias das contagens de sequências mapeadas.	52
Figura 5.6. Número de genes de acordo com suas médias das contagens de sequências mapeadas.	52
Figura 5.7. Comparação entre as contagens normalizadas e filtradas das réplicas biológicas.	53
Figura 5.8. " <i>MA Plot</i> " comparando os dados brutos entre as réplicas para cada amostra temporal.	54
Figura 5.9. " <i>MA Plot</i> " comparando os dados normalizados e filtrados.	55
Figura 5.10. Comparações dos valores de decaimento (coeficientes angulares ou <i>slopes</i>) entre os 4 tipos de mapeamento.	56
Figura 5.11. Comparações da média de <i>reads</i> mapeados entre os 4 tipos de tamanhos dos genes.	57
Figura 5.12. Comparação do número de contagens pré-normalização entre mapeamentos.	57
Figura 5.13. Diagrama de Venn da confiança da expressão diferencial em diferentes mapeamentos.	59

Figura 5.14. Diagrama de Venn dos genes diferencialmente expressos nos 4 mapeamentos.	60
Figura 5.15. “ <i>Smear Plot</i> ” da expressão diferencial entre os pontos 0 e 5 minutos. .	61
Figura 5.16. Clusterizações hierárquicas dos genes diferencialmente expressos. ...	63
Figura 5.17. Decaimento dos transcritos mais instáveis na série temporal.....	66
Figura 5.18. Decaimento dos transcritos mais estáveis na série temporal.	67
Figura 5.19. Histograma das taxas de decaimento para mapeamento “UTR”.	68
Figura 5.20. Padrão de expressão no ciclo de vida dos 500 transcritos mais estáveis.	73
Figura 5.21. Padrão de expressão no ciclo de vida de alguns genes de decaimento acentuado.	73
Figura 5.22. Comparação entre o tamanho do gene e a velocidade de degradação de seu transcrito.	74
Figura 5.23. Representatividade dos termos “GO” no grupo dos 500 transcritos mais estáveis.	75
Figura 5.24. Relação entre a reprodutibilidade do padrão de expressão dos genes entre as réplicas da série temporal e outras características dos genes.	76

LISTA DE TABELAS

Tabela 5.1. Contagem de sequências obtidas para cada biblioteca.	47
Tabela 5.2. Número e porcentagem das contagens modificadas após a “linha de base”.	51
Tabela 5.3. Número de genes diferencialmente expressos para vários limites de FDR.	58
Tabela 5.4. Os 20 genes mais confiáveis com expressão diferencial entre os pontos 0’ e 5’.	62
Tabela 5.5. Transcritos mais estáveis, independente do tipo de mapeamento.	69
Tabela 5.6. Transcritos menos estáveis, independente do tipo de mapeamento.	70

LISTA DE EQUAÇÕES

Equação 4.1. Equação da reta de regressão linear " $f(x)$ ".	41
Equação 4.2. Coeficiente de Variação dos Resíduos (CVR).	41
Equação 4.3. Coeficiente de correlação de Pearson (" r ").	42
Equação 4.4. Métrica de similaridade do método de clusterização de Eisen <i>et al.</i> (1998).	43

LISTA DE UNIDADES DE MEDIDA

bp – Pares de base (<i>base pairs</i> – unidade de comprimento de ácido nucléico)
°C – Grau Celsius (unidade de temperatura)
Da – Dalton (unidade de massa atômica, 1/12 da massa de um átomo isolado de Carbono 12 (^{12}C), aproximadamente a massa de um átomo de hidrogênio)
g – grama (unidade de medida de massa)
kb – Quilobase; comprimento de ácidos nucléicos, 1.000 bp
L – Litro (unidade de volume)
M – Molar (mols por Litro, unidade de concentração de soluto)
Mb – Megabases, 1.000 kb
pH – Potencial Hidrogeniônico (grandeza físico-química relativa à acidez-basicidade de uma solução)
U – Unidade (Unidades de Enzima)

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DA LITERATURA.....	14
2.1	O MODELO BIOLÓGICO <i>TRYPANOSOMA CRUZI</i>	14
2.2	A DOENÇA DE CHAGAS	16
2.3	O CICLO DE VIDA DO <i>T. CRUZI</i>	18
2.4	REGULAÇÃO DA EXPRESSÃO GÊNICA EM EUCARIOTOS	21
2.5	BIOLOGIA MOLECULAR E REGULAÇÃO DA EXPRESSÃO GÊNICA EM <i>T. CRUZI</i> 24	
2.6	TRANSCRIPTÔMICA E MEIA-VIDA DE RNAS	28
3	OBJETIVOS	33
3.1	OBJETIVO GERAL.....	33
3.2	OBJETIVOS ESPECÍFICOS.....	33
4	MATERIAIS E MÉTODOS.....	34
4.1	CULTIVO DE PARASITOS.....	34
4.2	INIBIÇÃO DA TRANSCRIÇÃO	34
4.3	ISOLAMENTO DO MRNA TOTAL	35
4.4	PREPARO DO MATERIAL PARA A REAÇÃO DE SEQUENCIAMENTO	35
4.5	REAÇÃO DE AMPLIFICAÇÃO EM EMULSÃO.....	37
4.6	PREPARO DAS LÂMINAS E REAÇÃO DE SEQUENCIAMENTO.....	37
4.7	MAPEAMENTO DAS SEQUÊNCIAS.....	38
4.8	NORMALIZAÇÃO E FILTRAGENS DOS DADOS	40
4.9	IDENTIFICAÇÃO DO PADRÃO DE DECAIMENTO	41
4.10	CLUSTERIZAÇÃO.....	42
4.11	CATEGORIZAÇÃO FUNCIONAL	43
5	RESULTADOS E DISCUSSÃO.....	44
5.1	QUALIDADE DAS AMOSTRAS DE MRNA	44
5.2	QUALIDADE DO SEQUENCIAMENTO	45
5.3	QUALIDADE DAS SEQUÊNCIAS	45
5.4	ANÁLISE DE FATORES INFLUENCIANDO A MAPEABILIDADE	47
5.5	EFICIÊNCIA DO PROCESSO DE PURIFICAÇÃO DE MRNA.....	50

5.6	QUANTIFICAÇÃO DE ELEMENTOS COM BAIXA CONTAGEM.....	50
5.7	ANÁLISE DA REPRODUTIBILIDADE BIOLÓGICA	53
5.8	ANÁLISE COMPARATIVA DOS MÉTODOS DE DEFINIÇÃO DA REGIÃO MAPEÁVEL	55
5.9	INFLUÊNCIA DA SINEFUNGINA	58
5.10	ANÁLISE DO DECAIMENTO	64
5.11	DESCRIÇÕES FUNCIONAIS DOS GENES MAIS ESTÁVEIS E INSTÁVEIS.	68
5.12	ANÁLISE DA REPRODUTIBILIDADE DO PADRÃO DE DECAIMENTO	75
6	CONCLUSÃO	77
7	PERSPECTIVAS	78
	REFERÊNCIAS BIBLIOGRÁFICAS	80

1 INTRODUÇÃO

Os sistemas biológicos são complexos e possuem características derivadas da interação de suas partes constituintes. A essas características é dado o nome de propriedades emergentes (MAYR, 1998), pois surgem da complexidade de interações inerentes ao sistema. Uma das propriedades emergentes mais conhecidas de sistemas considerados vivos é a homeostase. Ela surge da integração de mecanismos com função de manter os valores de variáveis bióticas e abióticas em um equilíbrio dinâmico. Basicamente, é criado um autocontrole do meio interno ao sistema biológico, no qual a estabilidade das variáveis tende a ser mantida. A capacidade de resposta de um organismo perante as modificações que ocorrem em seu contexto ambiental depende da agilidade e acurácia em que é realizado o controle da homeostase (ALBERTS *et al.*, 2004; VIDAL, 2009).

O controle de quais genes estão expressos em um dado momento é um dos principais fatores determinantes da habilidade de resposta da célula a um estímulo interno ou externo. Por esse motivo, estudos desse ajuste fino da expressão gênica são importantes para o entendimento da fisiologia celular. Para genes codificadores de proteínas esse controle pode ser classificado em 3 etapas principais que se encontram em série: transcricional, até o momento em que a RNA polimerase inicia a síntese do RNA (ácido ribonucléico, sigla derivada do termo em inglês *ribonucleic acid*) no núcleo celular; pós-transcricional, incluindo-se etapas ditas cotranscricionais e o controle de início da tradução; e pós-traducional, juntamente dos controles cotraducionais. Todas são observadas em eucariotos e são interdependentes (ALBERTS *et al.*, 2004; KEENE, 2007).

Para que o organismo reaja rapidamente a variações ambientais, alguns transcritos devem ser removidos da população total rapidamente e outros devem se manter estáveis ou aumentar sua proporção (BERNSTEIN *et al.*, 2002; ALBERTS *et al.*, 2004; KEENE, 2007). Um bom modelo biológico para estudos de decaimento de transcritos deve possuir diversos mecanismos de controle da expressão de seus genes por vias pós-transcricionais e pós-traducionais, com pouca regulação fina da etapa transcricional. O protozoário parasita *Trypanosoma cruzi* (responsável por causar doença de Chagas) regula sua expressão gênica majoritariamente em etapas

de controle após a maturação de seus mRNAs (RNAs mensageiros, sigla do inglês *messenger RNAs*). Esta é feita por um processo acoplado de *trans-splicing* e poliadenilação de transcritos primários policistrônicos (VANHAMME & PAYS, 1995; CAMPBELL *et al.*, 2003; DOSSIN & SCHENKMAN, 2005; SIMPSON *et al.*, 2006; CLAYTON & SHAPIRA, 2007).

O Instituto Carlos Chagas (ICC – Fiocruz - região Sul), local onde foi realizado este trabalho, possui um projeto intitulado “Projeto Reguloma de *T. cruzi*” que tem como objetivo final integrar os dados obtidos experimentalmente e identificar os módulos de correção gênica e seus elementos regulatórios em *cis* (por exemplo, motivos de sequências e estruturas secundárias de RNA) e em *trans* (como proteínas ligadoras de ácidos nucleicos). Os experimentos realizados focam na análise do transcriptoma (conjunto total de mRNAs celulares) mediante submissão do parasito a alterações de variáveis ambientais (como estresse nutricional, de temperatura e oxidativo) e a presença de drogas de diversos tipos (inibidores de vias de biossíntese e fármacos tripanocidas).

Em meu projeto de mestrado, que faz parte dos experimentos do projeto Reguloma, foi avaliado o decaimento do transcriptoma de epimastigotas da cepa Y de *T. cruzi*. Sendo assim, foram realizados estudos sobre a etapa pós-transcricional de controle da expressão gênica. As análises feitas focaram no estudo do decaimento dos mRNAs e suas possíveis implicações fisiológicas.

2 REVISÃO DA LITERATURA

2.1 O MODELO BIOLÓGICO *Trypanosoma cruzi*

Dos cerca de 60 grupos parafiléticos distintos de protistas, 27 são flagelados (PATTERSON & LARSEN, 1991). Pertencendo a esse grupo, o *Trypanosoma* (*Schizotrypanum*) *cruzi* (Chagas, 1909) é um eucarioto unicelular do Super-reino Metakaryota, Reino Protozoa, pertencente ao Infra-reino e Filo Euglenozoa, Classe

Kinetoplastea, Ordem Kinetoplastida (HONIGBERG, 1963; CAVALIER-SMITH, 1993; CAVALIER-SMITH, 2010). Essa Ordem possui as Famílias Bodonidae e Trypanosomatidae, sendo que na primeira estão os gêneros *Bodo*, *Cruzella* e *Neobodo*, que são geralmente de vida livre, e a segunda abrange os gêneros *Angomonas*, *Crithidia*, *Leptomonas*, *Phytomonas*, *Trypanosoma* e *Leishmania*, que são todos parasitos mono ou heteroxênicos (GODOI, 2000; SIMPSON *et al.*, 2006; TEIXEIRA *et al.*, 2011). Uma característica específica dos organismos da Classe Kinetoplastea é a existência de pelo menos uma região especializada de sua mitocôndria única chamada cinetoplasto que aloja uma rede de DNA (ácido desoxirribonucleico, sigla do inglês *deoxyribonucleic acid*) organizada, denominada DNA do cinetoplasto ou kDNA, em forma de bastão, mas que pode ter sua forma e organização interna alterada dependendo da fisiologia do protozoário (CAVALIER-SMITH, 2010). Além do cinetoplasto, outras estruturas peculiares estão presentes, como os peroxissomos, reservossomos, microtúbulos subpeliculares, glicossomos e um ou dois flagelos estão presentes. O gênero *Trypanosoma* apresenta cinco subgêneros descritos. É no subgênero *Schizotrypanum* que está classificado o *Trypanosoma (Schizotrypanum) cruzi*, foco deste trabalho (CAVALIER-SMITH, 1993; DE SOUZA, 2000).

Juntamente com o gênero *Leishmania*, as espécies de *Trypanosoma* são parasitos que possuem dois hospedeiros, um artrópode e outro mamífero, com impacto humano, seja causando doenças humanas ou de animais domesticados. Os parasitos do gênero *Leishmania* causam a leishmaniose cutânea, de mucosas e visceral, em humanos e outros mamíferos como o cão (CHAPPUIS *et al.*, 2007), enquanto alguns membros do gênero *Trypanosoma* podem causar, em humanos, a doença do sono ou tripanossomíase africana, causada por *T. brucei gambiense* e *T. brucei rhodesiense* (GARCIA *et al.*, 2006) e a doença de Chagas ou tripanossomíase americana, causada pelo *T. cruzi*, causando grande morbidade e mortalidade. Outras espécies do gênero *Trypanosoma* são agentes etiológicos de doenças em ruminantes e equinos gerando grandes perdas econômicas e sofrimento animal (JONES & DÁVILA, 2001; GILLINGWATER *et al.*, 2007).

O *T. cruzi* é uma espécie com grande diversidade intraespecífica, originada por divisão clonal e subsequente evolução paralela das diferentes linhagens e cepas. As várias cepas apresentam diferenças quanto à sua capacidade infectiva em seus vários hospedeiros. É observado, também, variação em virulência e

resistência a agentes líticos derivados de triatomíneos (ZINGALES *et al.*, 1999). Estudos de filogenia desse protozoário mostram que sua população pode ser dividida ao menos em dois grandes grupos, denominados *T. cruzi* I, geralmente mais relacionado ao ciclo silvático (ou silvestre) de transmissão, e *T. cruzi* II, relacionado em maior frequência ao ciclo domiciliar (também usado o termo doméstico). Este, por sua vez, possui cinco subgrupos: IIa, IIb, IIc, IId e IIe (BRISSE *et al.*, 2000; EL-SAYED *et al.*, 2005a). Posteriormente, foi utilizada a nomenclatura DTU (do inglês, *Discrete Typing Unit*) I e II ao invés de *T. cruzi* I e II (TIBAYRENC, 2003). Uma nova proposta de nomenclatura classifica as cepas desse parasito em seis subpopulações denominadas *T. cruzi* I a VI, de acordo, respectivamente, com as DTUs I, IIb, IIc, IIa, IId e IIe (ZINGALES *et al.*, 2009). A distância filogenética entre as linhagens de *T. cruzi* é tão grande quanto a distância entre as espécies do gênero *Leishmania* (dixênico) e *Crithidia* (monoxênico), que possuem ciclos de vida bastante diferentes (JUNQUEIRA *et al.*, 2005). Por esse motivo, alguns autores sugerem que essa espécie seja considerada um complexo ou que ocorra uma divisão em espécies ou subespécies distintas (DEVERA *et al.*, 2003; JUNQUEIRA *et al.*, 2005).

2.2 A DOENÇA DE CHAGAS

Em 1909, o cientista brasileiro Carlos J. Ribeiro das Chagas descreveu uma doença causada por um protozoário que afetava os humanos. Descreveu, também, o vetor, o ciclo de vida e a forma de transmissão do parasito (CHAGAS, 1909). A doença causava um quadro grave e, em sua homenagem, foi denominada de doença de Chagas. Essa doença apresenta uma área de incidência por toda a América Latina e é um grave problema de saúde pública, afetando de 16 a 18 milhões de pessoas atualmente, com mais de 100 milhões em risco de infecção (BARRETT *et al.*, 2003; VILLELA *et al.*, 2007; DUMONTEIL, 2007). Aproximadamente 45 mil indivíduos morrem anualmente devido às complicações, principalmente cardíacas e/ou digestórias, geradas pela doença (TDR, 2003).

Essa enfermidade pode ter diversas apresentações, não sendo raros os casos assintomáticos. Em sua fase crônica, 30 a 40% dos infectados desenvolve o

quadro denominado cardiomiopatia chagásica crônica, que causa morte por insuficiência cardíaca, enquanto os outros são assintomáticos ou possuem apenas sintomas como astenia (fraqueza muscular ao acordar, derivada de processos infecciosos) (TDR, 2003; TEIXEIRA *et al.*, 2006; DUMONTEIL, 2007). A doença de Chagas no Brasil possui alto índice de prevalência, principalmente em Minas Gerais (VILLELA *et al.*, 2007), estado onde a doença foi descoberta (CHAGAS, 1909). A terapia utilizada é feita com nitrofuranos ou nitroimidazol que, além de gerar muitos efeitos colaterais, é pouco efetiva, sendo que algumas linhagens de *T. cruzi* apresentam resistência a essas drogas (DUMONTEIL, 2007).

O parasito persiste indefinidamente, mesmo em níveis mínimos, assim como a resposta imunológica contra o mesmo, resultando em acúmulo de lesões teciduais que acabam por gerar, em alguns casos, alterações morfológicas e funcionais dos tecidos afetados. O comportamento do sistema imune sobre as diferentes formas celulares do parasito é distinto (COURA & DE CASTRO, 2002; ANDRADE & ANDREWS, 2005).

A doença em questão é transmitida por insetos da Ordem Hemiptera, da Família hematófaga Reduviidae, Subfamília Triatominae. Os principais gêneros de barbeiros envolvidos no ciclo de transmissão do parasito são *Triatoma*, *Panstrongylus* e *Rhodnius*. Os triatomíneos picam os mamíferos selvagens, reservatórios naturais do *T. cruzi*, para alimentar-se de seu sangue (NEVES, 1991). Com a destruição sistemática de seus habitats naturais e pelo processo de colonização, os vetores do *T. cruzi* entraram em maior contato com os humanos, propiciando a infecção (DIAS, 2000). A infecção humana pelo protozoário se dá, principalmente, (i) pela penetração deste presente nas fezes do hemíptero em mucosas ou lesões da pele, como a lesão causada pela picada do vetor, durante o repasto sanguíneo, (ii) por contaminação em transfusão de sangue, (iii) congenitamente ou (iv) por ingestão de triatomíneos, fatos isolados que foram associados a caldo de cana e suco de açaí (NEVES, 1991; TEIXEIRA *et al.*, 2006). A vigilância sanitária nacional, atualmente, tem agido de forma bastante eficaz para evitar a contaminação por transfusão sanguínea, diminuindo drasticamente novos casos da doença por essa via de infecção. No entanto, aumentaram-se os casos por infecção oral (COURA, 2006; COURA, 2007).

Após a adaptação de algumas espécies de triatomíneos ao ambiente fornecido pela presença humana, dois ciclos distintos, porém relacionados, foram

estabelecidos, o doméstico e o silvestre (FIGURA 2.1). O tipo de domicílio em que os insetos se estabeleceram em seu ciclo doméstico foram os de barro e pau-a-pique de zona rural, bem como chiqueiros e galinheiros. A presença constante e localizada de animais domésticos e do homem facilitou grandemente o processo de adaptação do inseto vetor por fornecer-lhe alimento. No ambiente doméstico, existe também maior proteção contra seus predadores naturais. Aliado a isso, as frestas das paredes de barro seco facilitam a reprodução do vetor da doença (NEVES, 1991).

2.3 O CICLO DE VIDA DO *T. cruzi*

Em seu ciclo de vida, o *T. cruzi* sofre alterações seriais em sua morfologia celular. O *T. cruzi* alterna também seu ambiente: um hospedeiro intermediário artrópode, o inseto barbeiro, e um mamífero definitivo, que são animais de grande distância filogenética, cujo ancestral comum viveu entre 550 a 800 milhões de anos atrás (KNOLL & CARROLL, 1999; TRAVIS, 2007).

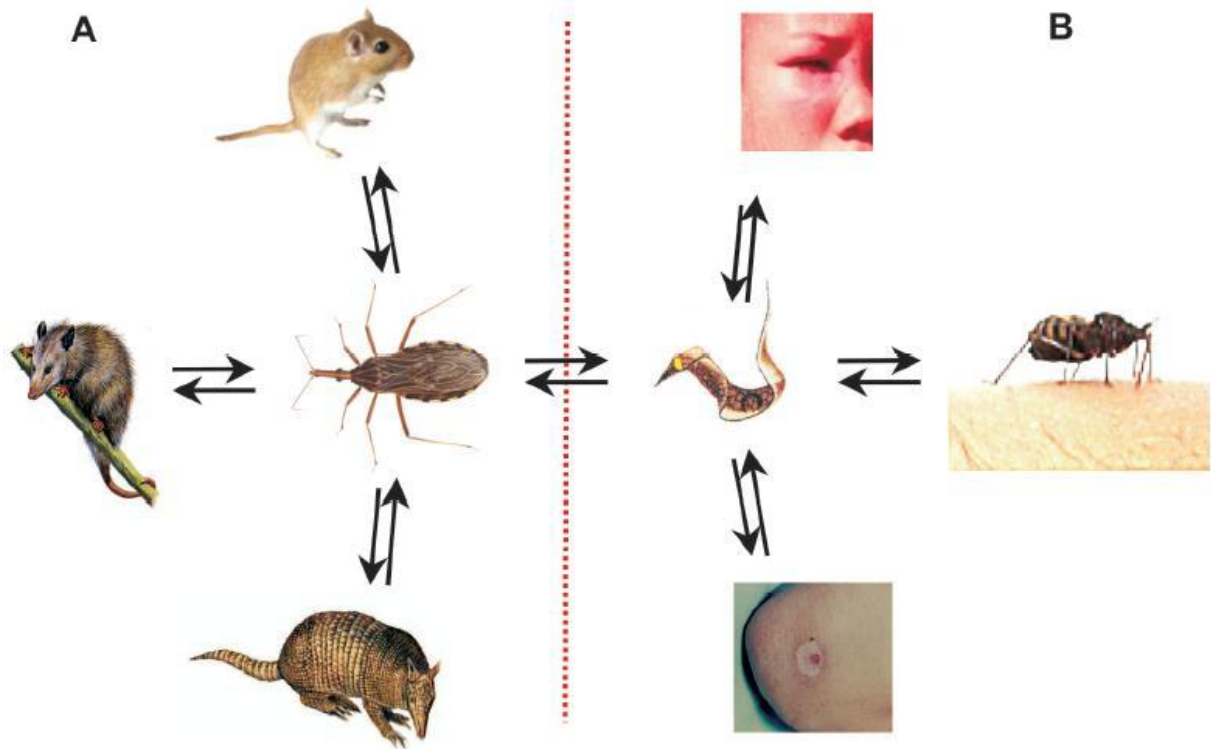


FIGURA 2.1. CICLOS SILVESTRE (A) E DOMÉSTICO (B) DOS TRIATOMÍNEOS E DO PROTOZOÁRIO *T. CRUZI*.

Em A, podem ser observados os mamíferos selvagens como um roedor (acima), um marsupial (à esquerda) e um xenartro (abaixo), que são picados pelo triatomíneo (à direita). Em B têm-se um “chagoma” (abaixo) e um sinal de Romaña (acima), que são marcas da picada (à direita) do vetor com posterior infecção pelo *T. cruzi* (à esquerda).

FONTE: TEIXEIRA *et al.*, 2006.

Resumidamente, o ciclo se inicia com a contaminação do triatomíneo pelo tripanossomo durante sua alimentação com o sangue do hospedeiro mamífero infectado. O hemíptero, no repasto, suga também formas celulares do *T. cruzi* que são infectivas, mas não replicativas, chamadas tripomastigotas sanguíneos. Estes possuem uma forma alongada, flagelo associado a uma membrana ondulante e cinetoplasto em posição posterior em relação ao núcleo. Dentro do intestino do barbeiro, os tripomastigotas em questão diferenciam-se para outra forma celular chamada epimastigota, também alongada, mas não infectiva e sim replicativa, com o cinetoplasto em posição anterior (região mais próxima da extremidade final do flagelo) em relação ao núcleo. Os epimastigotas, ao alcançarem a porção final do trato digestório do inseto, se diferenciam na forma celular tripomastigota metacíclico que não replica, mas que pode infectar mamíferos quando as excretas do hemíptero que contém tais formas entram em contato com mucosas ou lesões na epiderme causadas pelo inseto, pela perfuração da mesma, ao se alimentar. Ao entrarem no

mamífero, os tripomastigotas metacíclicos infectam células de diversos tipos e transformam-se em amastigotas, formas sem flagelo exteriorizado que replicam intracelularmente até saturarem o citoplasma hospedeiro e modificar sua forma para tripomastigotas sanguíneos. Estes, ao se movimentarem fortemente, rompem a célula hospedeira e são liberados na corrente sanguínea, podendo infectar novas células. Dessa forma, o ciclo se completa quando o vetor realiza o repasto e se infecta novamente com tripomastigotas circulantes (NEVES, 1991; TYLER & ENGMAN, 2001; MACEDO *et al.*, 2002). As etapas do ciclo podem ser vistas na FIGURA 2.2 logo abaixo.

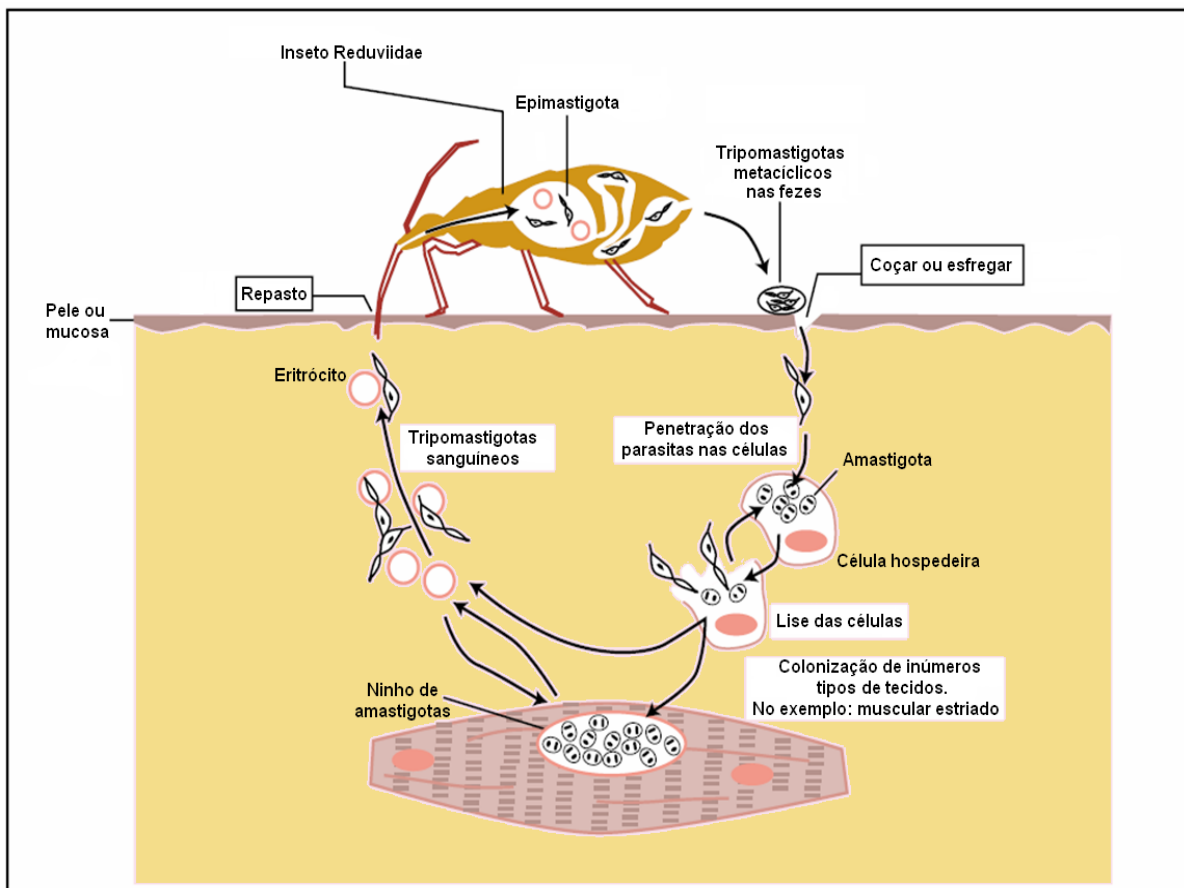


FIGURA 2.2. O CICLO DE VIDA DE *T. CRUZI* (SIMPLIFICADO).

FONTE: modificado de MACEDO *et al.*, 2002.

Uma etapa do ciclo de vida do *T. cruzi* é a transformação de epimastigota para tripomastigota metacíclico, denominada metaciclogênese. Esta pode ser realizada *in vitro* sem a utilização de células hospedeiras (CONTRERAS *et al.*, 1985). Esse processo gera uma forma infectiva passível de utilização para ensaios

in vivo de infectividade e é um modelo apropriado para o estudo dos mecanismos envolvidos na regulação da diferenciação celular e da expressão gênica no parasito (CONTRERAS *et al.*, 1988; BONALDO *et al.*, 1988). A metaciclogênese em *T. cruzi* pode ter uma grande relação com o decaimento de moléculas de RNA na célula pelo fato de que a transcrição, de forma geral, se encontra bastante diminuída na forma celular não replicativa que surge ao final desse processo sofrido pelo parasito (ELIAS *et al.*, 2001; FERREIRA *et al.*, 2008). A mesma queda nos níveis transcricionais parece ocorrer nas formas tripomastigotas de *T. brucei* (EHLERS *et al.*, 1987; PAYS *et al.*, 1993 *apud* CLAYTON & SHAPIRA, 2007). Adicionalmente, a meia-vida de mRNAs em formas sanguíneas (infectivas e não replicativas) é menor (para vários transcritos) que em formas procíclicas (replicativas) em *T. brucei*, possivelmente pela diferença de taxa de divisão celular e temperatura em que essas formas vivem (EHLERS *et al.*, 1987, CLAYTON & SHAPIRA, 2007). Sendo assim, a estabilidade *versus* a degradação de transcritos pode fazer parte dos gatilhos que disparam os eventos de diferenciação celular.

2.4 REGULAÇÃO DA EXPRESSÃO GÊNICA EM EUCARIOTOS

Todos os seres vivos celulares devem ajustar de forma fina a expressão de seus genes para que sobrevivam perante as mudanças que naturalmente ocorrem em todos os tipos de ambientes. Para isso, o processo de evolução gerou pressão seletiva para que surgissem, e fossem mantidos, diversos mecanismos de controle.

Nos eucariotos mais utilizados como modelo de estudos (por exemplo, espécies dos gêneros *Saccharomyces*, *Caenorhabditis*, *Drosophila*, *Xenopus*, *Mus*, *Homo* e *Arabidopsis*) o principal controle que define a ativação ou repressão dos genes se dá na etapa da transcrição, na qual fatores gerais de transcrição ligam-se a sítios denominados promotores, marcando o sítio de início da transcrição para a RNA polimerase (CLAYTON, 2002; CLAYTON & SHAPIRA, 2007; ALBERTS *et al.*, 2004). Dessa forma é sintetizado o pré-mRNA ou transcrito primário, que sofre clivagens pelo processo de *cis-splicing* para remoção de íntrons e modificações nas

duas terminações da molécula, com adição do *cap* 5' e a cauda poli-A a 3' (ALBERTS *et al.*, 2004).

Na extremidade 5' ocorre a adição de um nucleotídeo modificado 7-metil-guanosina (*cap* 5') por ligação do tipo 5' \rightarrow 5', em um processo denominado “capeamento” de mRNA. O *cap* 5' tem como função principal proteger essa extremidade da degradação por exorribonucleases e atuar como um sinal de que esta porção da molécula está madura e pronta para servir de sítio de ligação para proteínas ligadoras de *cap* 5', passo necessário para a posterior exportação nuclear.

Já na outra extremidade, a região 3', ocorre uma clivagem endonucleolítica e em sequência a síntese de uma cauda de adeninas realizada pela enzima poli-A polimerase. Essa modificação se dá em um sítio marcado pelo dinucleotídeo CA nas proximidades (de 10 a 30 nucleotídeos) a jusante de sequências conservadas cujo consenso é AAUAAA. A cauda poli-A sintetizada possui tamanho variado, em média com cerca de 200 nucleotídeos nos organismos modelo. Assim como o *cap* 5', serve para proteção da molécula contra sua degradação por exonucleases e também funciona como sítio de ligação para proteínas (proteínas ligadoras de poli-A) que sinalizam a maturação do terminal 3' e auxiliam na exportação da molécula para o citosol (ALBERTS *et al.*, 2004).

Existe no núcleo celular um mecanismo de vigilância que detecta mRNAs danificados por clivagem ou com erros de codificação. Por exemplo, um erro gerado durante a transcrição, como a falta de um evento de *cis-splicing* manter um íntron indesejado no mRNA maduro e, por consequência, existir um códon de parada da síntese proteica antes do verdadeiro sinal de término em um éxon a jusante. A esse tipo específico de controle de qualidade, que verifica sinais de parada internos à CDS (região codificadora de proteína), dá-se o nome de decaimento mediado por mutação sem sentido (sigla NMD, do inglês *nonsense-mediated decay*) (HENTZE & KULOZIK, 1999). Diversas proteínas ligadoras de RNA (RBPs, do inglês *RNA Binding Proteins*) e vários complexos proteicos e de ribonucleoproteínas são responsáveis por interagir e sinalizar para a degradação espécies de macromoléculas (tanto proteínas quanto RNAs) incorretas e/ou perigosas para a célula. Parte da mesma maquinaria utilizada para o controle por NMD está envolvida na destruição de íntrons excisados e outros fragmentos derivados de diversos processamentos de RNAs. Também foi observado que esse mecanismo está

envolvido com a regulação da expressão gênica (HENTZE & KULOZIK, 1999; ALONSO, 2005; CHANG *et al.*, 2007; KEENE, 2007; GLISOVIC *et al.*, 2008).

Além da ativação da transcrição de um gene e, em sequência, de seu produto gênico, o controle de quais genes têm sua expressão desativada é crucial para a homeostase e a resposta ao ambiente. Portanto, a abundância de mRNAs de um determinado gene na célula é proporcional à sua taxa de transcrição e inversamente proporcional a sua taxa de degradação (CLAYTON, 2002; CLAYTON & SHAPIRA, 2007; KEENE, 2007; SUNNERHAGEN, 2007).

Existem duas vias principais de degradação de mRNAs que é conservada na maioria dos eucariotos estudados. Uma degrada o mRNA iniciando por retirar o *cap* 5' (descapamento) do transcrito e posteriormente exorribonucleases que agem em sentido 5' para 3' clivam a molécula em nucleosídeos mono ou difosfatados (dependendo da forma de catálise da exonuclease). A outra via segue em sentido inverso na molécula de RNA, por enzimas que degradam RNA da extremidade 3' para a 5', que independe do descapamento (DECKER, 1998; CLAYTON, 2002; HOUSELEY *et al.*, 2006).

Algumas destas ribonucleases conhecidas fazem parte de um complexo denominado exossomo, descoberto em 1997 por Mitchell e colaboradores, em leveduras de brotamento (MITCHELL *et al.*, 1997). O exossomo está envolvido no decaimento de moléculas de mRNA no citoplasma e, no núcleo, na maturação do rRNA (*ribosomal RNA*, RNA ribossômico) 5.8S e na degradação de intermediários de processos como *splicing* e controle de qualidade de mRNAs antes de sua exportação nuclear (MITCHELL *et al.*, 1997; DECKER, 1998; HOUSELEY *et al.*, 2006; CLAYTON & SHAPIRA, 2007). Esse complexo já foi identificado em *Trypanosoma brucei*, apesar de ligeiramente diferente do complexo de exorribonucleases de leveduras com relação às subunidades que o compõem (ESTÉVEZ *et al.*, 2001; CLAYTON & SHAPIRA, 2007). Ambas dependem da degradação inicial da cauda poli-A dos transcritos maduros por enzimas poli-A ribonucleases (DECKER, 1998; CLAYTON, 2002; CLAYTON & SHAPIRA, 2007). Posteriormente, uma das vias, ou ambas concomitantemente, é ativada e culmina na destruição dos mRNAs-alvo.

As RBPs são proteínas de enorme importância em inúmeros mecanismos de regulação da expressão gênica por atuarem em diversos complexos que interagem com RNAs (MARIS *et al.*, 2005; KEENE, 2007). Essas proteínas estão claramente

envolvidas em ligação a regiões não traduzidas (UTRs, do inglês *untranslated region*) dos mRNAs que sinalizam para eventos de processamento, transporte e endereçamento, estocagem e inclusive estabilidade *versus* degradação dos transcritos maduros (KEENE, 2007).

É importante salientar que as etapas de controle da expressão gênica são diversificadas e em grande número. Nesse contexto o evento de degradação de mRNAs se contextualiza como sendo o somatório final de todo esse conjunto de controle da expressão gênica (EHLERS *et al.*, 1987; MITCHELL *et al.*, 1997; ESTÉVEZ *et al.*, 2001; CLAYTON, 2002; CLAYTON & SHAPIRA, 2007).

2.5 BIOLOGIA MOLECULAR E REGULAÇÃO DA EXPRESSÃO GÊNICA EM *T. cruzi*

Diferentemente da maioria dos eucariotos conhecidos, os tripanossomatídeos possuem transcrição policistrônica, como observado em bactérias, porém sem a organização característica de operons, isto é, os genes constituintes de uma unidade de transcrição policistrônica não são relacionados entre si em uma mesma via. Corroborando tais afirmações, foram observadas diferenças de expressão para genes de vias metabólicas diferentes que fazem parte de um mesmo transcrito primário, mas que não são vistos corregulados (VANHAMME & PAYS, 1995; CAMPBELL *et al.*, 2003; SIMPSON *et al.*, 2006).

Uma outra característica marcante da Classe Kinetoplastea é o *trans-splicing* dos seus RNAs mensageiros. Esse processo consiste na transferência de um fragmento de molécula de RNA, denominada sequência *spliced-leader* (SL) ou mini-éxon, já contendo o *cap* 5', formado por um nucleotídeo 7-metilguanosina trifosfatado, à porção 5' de cada cístron. Ocorre de forma concomitante à poliadenilação na região 3' de cada unidade codificadora. O *T. cruzi* também difere da maioria dos eucariotos mais estudados quanto à presença de íntrons, pois em seu genoma são raros os genes com sequências intrônicas, como é o caso do gene da enzima poli-A polimerase, removido por *cis-splicing* (MAIR *et al.*, 2000; DJIKENGI *et al.*, 2001; PALFI *et al.*, 2002; CAMPBELL *et al.*, 2003; SIMPSON *et al.*, 2006).

De maneira geral, os tripanossomatídeos não apresentam regulação transcricional gene-específica, sendo que a regulação pós-transcricional da expressão gênica, e da síntese protéica, e as modificações e controle pós-traducionais são muito importantes (VANHAMME & PAYS, 1995; CAMPBELL *et al.*, 2003; SIMPSON *et al.*, 2006).

Em julho de 2005 foi publicada a primeira versão do sequenciamento do genoma nuclear do *Trypanosoma cruzi*. A cepa selecionada para o sequenciamento (CL Brener) está dentre as mais estudadas linhagens de *T. cruzi*. Ela é resultado de uma hibridação de clones dos subgrupos IIb e IIc, e se inclui no subgrupo IIe do grupo *T. cruzi* II. Já era conhecida na época a alta heterozigosidade do protozoário e o fato de que vários cromossomos homólogos possuíam tamanhos diferentes (EL-SAYED *et al.*, 2005a). Esses fatos, associados à quantidade de família gênicas com grande similaridade de sequência entre si, dificultaram a montagem do genoma de *T. cruzi*.

Além do genoma nuclear, o *T. cruzi* possui um genoma mitocondrial bastante peculiar, característica comum aos Kinetoplastea. O cinetoplasto, uma região especializada no interior da mitocôndria única desse organismo, apresenta uma imbricada rede de DNA (o kDNA), que é composta por várias moléculas de DNA circular, categorizadas em dois tipos. Na primeira categoria, são encontradas moléculas circulares grandes (35-50 kb), análogas ao DNA mitocondrial dos outros eucariotos. Essas moléculas de DNA, chamadas maxicírculos, estão presentes em poucas dezenas de cópias e codificam mRNAs, rRNAs e RNAs guia (gRNAs, sigla derivada do inglês *guide RNAs*). Os gRNAs participam no processo de edição de mRNAs mitocondriais, que insere (e, com menos frequência, também retira) uridinas aos transcritos. O outro tipo de kDNA consiste de moléculas circulares pequenas, denominados minicírculos, medindo entre 0,6 e 5 kb, com dezenas de milhares de cópias por cinetoplasto e que possuem regiões conservadas intercaladas com regiões variáveis que podem ser utilizadas para caracterizações filogenéticas em tripanossomatídeos. Essa miríade de RNAs codificados pelos maxicírculos são transcritos por uma RNA polimerase mitocondrial (mtRNAPol), codificada no DNA nuclear e que possui apenas uma subunidade (BENNE, 1994; JUNQUEIRA *et al.*, 2005; CAMPBELL *et al.*, 2003; SIMPSON *et al.*, 2006). Além disso, os tRNAs (RNAs de transferência) utilizados na mitocôndria dos cinetoplastídeos localizam-se totalmente no genoma nuclear (CAMPBELL *et al.*, 2003).

Para o projeto genoma do parasito, foi utilizada a técnica de WGS (*Whole Genome Shotgun*), sendo que a quantidade de sequência gerada excede largamente o tamanho do genoma haplóide, o que é necessário para que se obtenha uma boa representatividade da maior parte do genoma. A quantidade de vezes que as sequências geradas excedem o tamanho do genoma haplóide é denominada cobertura e, no caso do *T. cruzi*, foi de 19 vezes. O tamanho do genoma haplóide foi calculado em cerca de 60 Mb, com aproximadamente 12 mil pares de alelos. Para o genoma diplóide foi estimado entre 106,4 e 110,7 Mb, no qual, em média, existe 5,4% de diferença entre os dois haplótipos, sendo menor (2,2%) a diferença dentro das porções codificadoras do genoma. Mais da metade do genoma é constituído de sequências repetitivas, principalmente de retrotransposons, RHSs (*retrotransposon hot spot proteins*) e proteínas de superfície, como as trans-sialidases, as mucinas, as glicoproteínas de 63 kDa (gp63), as DGF-1s (*dispersed gene family 1*) e as MASPs (*mucin-associated surface proteins*, proteínas de superfície associadas a mucinas, uma família nova descoberta com o projeto genoma desse protozoário). Finalmente, foram atribuídas funções putativas a 50,8% dos genes inicialmente preditos (EL-SAYED *et al.*, 2005a). Essa proporção relativamente baixa de predição funcional, associado ao fato de que muitas das predições são muito genéricas, reforça a necessidade de se caracterizar mais profundamente a função das proteínas de *T. cruzi*.

Concomitantemente com a publicação do genoma, foi realizada a caracterização proteômica do *T. cruzi* (ATWOOD III *et al.*, 2005). Os autores identificaram 2.784 proteínas como expressas em diferentes fases do ciclo do parasito, das quais 36% são proteínas hipotéticas, sendo que 348 delas não haviam sido preditas no genoma. Muitas não são estágio-específicas, mas ocorrem em maior quantidade em uma ou outra forma celular, como é o caso da maioria das trans-sialidases, que são altamente expressas em tripomastigotas e diminuem drasticamente sua presença em epimastigotas. Das proteínas encontradas no trabalho de proteômica, respectivamente, 49, 125 e 74 são específicas das fases amastigota, tripomastigota e epimastigota. Nenhuma proteína foi identificada como expressa unicamente em tripomastigotas metacíclicos. Por outro lado, 84 proteínas aparecem em todas as fases exceto em metacíclicos. O mesmo pode ser dito para amastigotas, tripomastigotas e epimastigotas para 93, 538 e 161 proteínas, respectivamente. Contudo, 838 proteínas foram localizadas em todos os estágios

analisados do ciclo vital do *T. cruzi* (ATWOOD III *et al.*, 2005). Dados como esses ressaltam a complexidade da regulação da expressão gênica nesse organismo e abrem caminho para análises mais profundas que podem vir a ser relacionadas à patologia e epidemiologia da doença.

É importante ressaltar que o estudo da especificidade da expressão protéica é importante para melhor entendimento sobre a forma de vida do organismo assim como sua biologia molecular, celular e epidemio-patológica, podendo levar ao desenvolvimento de drogas melhores, menos tóxicas, mais específicas e efetivas, para o tratamento, bem como formas mais eficazes de prevenção da doença. Essa especificidade de expressão pode estar intimamente relacionada com a estabilidade de seus mRNAs (ELIAS *et al.*, 2001; GENTIL *et al.*, 2009).

Além da proteômica, outra metodologia, o microarranjo de DNA, mostrou-se muito útil para o entendimento da regulação da expressão gênica, visto que possibilita a análise simultânea de quais genes de todo um genoma são transcritos (SCHENA *et al.*, 1995; LANDER, 1999). Avaliar o transcriptoma (conjunto de RNAs que foram transcritos, a partir de seus genes, em um dado momento) de uma célula fornece informações importantes para esclarecer melhor suas características funcionais e até mesmo sua morfologia, pois permite descrever quais genes estão sendo transcritos e em que condições (SCHULZE & DOWNWARD, 2001).

Em *T. cruzi* já foram realizados alguns trabalhos que utilizaram essa metodologia para avaliar expressão diferencial de genes. Tanto estudos sobre expressão estágio específica (MINNING *et al.*, 2009) quanto sobre a resposta do transcriptoma do parasito, ao ser exposto a algum tratamento como, por exemplo, radiação gama (GRYNBERG *et al.*, 2012) e drogas como inibidores da síntese de esteróis (KESSLER, 2010). Esse mesmo método de análise transcriptômica já foi usado para observar a resposta de células hospedeiras ao serem infectadas pelo parasito (VAENA DE AVALOS *et al.*, 2002; GARG *et al.*, 2003). Além de avaliar a expressão gênica ao nível do transcriptoma, os microarranjos também foram a base metodológica em experimentos sobre a variação do número de cópias gênicas e as aneuploidias dos cromossomos desse parasito (MINNING *et al.*, 2011) e foram utilizados para ensaios de ribonômica (interação proteína-RNA) de uma proteína (TcPUF6) ligadora de RNA (DALLAGIOVANNA *et al.*, 2008).

Sendo assim, nota-se que essas abordagens de estudos utilizando microarranjos de DNA, ao serem associadas a outras caracterizações em larga

escala (ou “ômicas”), como proteômica e interatômica, possibilitam um grande aumento na caracterização de processos biológicos (VIDAL, 2001; SCHULZE & DOWNWARD, 2001).

2.6 TRANSCRIPTÔMICA E MEIA-VIDA DE RNAS

Em meio à década de 1990, surgiram duas técnicas importantes para os estudos em larga escala dos transcritos celulares: *SAGE* (*Serial Analysis of Gene Expression*; VELCULESCU *et al.*, 1995) e microarranjos de DNA (SCHENA *et al.*, 1995; DeRISI *et al.* 1997; SPELLMAN *et al.*, 1998). Conforme já foi mencionado sobre os microarranjos, diversas técnicas se popularizaram e foram bastante utilizadas para a análise de transcriptoma e, dessa forma, tornaram-se metodologias com alto impacto científico. Mais recentemente, foi introduzida a técnica de sequenciamento de transcriptoma (*RNA-Seq Whole Transcriptome* ou somente *RNA-Seq*), que se mostrou mais eficiente e produtiva devido a diversos fatores, tais como: quantificar de maneira mais precisa o número de moléculas de RNA obtidas nas amostras, ter uma melhor cobertura genômica, devido ao fato de não ser necessário o conhecimento da sequência completa do genoma do organismo a ser estudado e possuir maior resolução e qualidade de sinal pela independência do uso de sonda (assim sendo, não existe hibridização cruzada e o ruído inerente à técnica é muito mais baixo) (SHENDURE & JI, 2008; WANG *et al.*, 2009).

Dos tripanossomatídeos, apenas em *Trypanosoma brucei* já foi publicado um trabalho utilizando *RNA-Seq* para compreender seu transcriptoma, porém, sem avaliações sobre o decaimento dos transcritos (SIEGEL *et al.*, 2010). Para construção de uma base de dados ampla relativa ao transcriptoma de *T. cruzi* é preciso avaliar diretamente qual é a dinâmica da meia-vida do conjunto total de mRNAs expressos, através do uso de inibidores de transcrição e avaliação em larga escala do ritmo de decaimento, o qual representa o resultado final de uma complexa rede de interações moleculares que determinam o perfil funcional dos mRNAs. A avaliação desse ritmo e a comparação das mudanças desses perfis de decaimento entre diferentes formas evolutivas de *T. cruzi* constitui uma fonte de dados valiosa

para a determinação da dinâmica de regulação dos níveis de mRNA em *T. cruzi*. Esses dados associados aos já existentes, relativos a outros sistemas e processos biológicos de *T. cruzi*, permitirão a identificação de grupos de genes coregulados de forma muito mais produtiva, possibilitando um melhor entendimento do processo de regulação da expressão gênica de *T. cruzi*.

Para avaliar decaimento de moléculas, deve ser mensurada a quantidade inicial de moléculas existentes na célula em um dado momento e, após um tempo determinado, ocorrer uma nova mensuração, considerando apenas as mesmas moléculas que antes foram contabilizadas. No caso de mRNAs, não podem ser computados mRNAs que foram sintetizados entre os tempos (inicial e final) de análise. Para tanto, é preciso marcar quais foram as moléculas de RNA que foram avaliadas inicialmente ou, de forma mais simples, inibir a transcrição de novos RNAs.

Um exemplo de inibidor do mecanismo de transcrição é a droga actinomicina D (FIGURA 2.3), que é um antibiótico e antineoplásico derivado de aminoácidos que foi descoberto em 1940 por Waksman e Woodruff em culturas de bactérias Gram-positivas da espécie *Actinomyces antibioticus* (KOBAYASHI & KONOPKA, 2005). Segundo Sobell (1985), esse composto atua através da formação de um complexo estável com DNA dupla-fita (FIGURA 2.4). O anel cromóforo (anel fenoxazona, que confere cor à droga) intercala entre dinucleotídeos pareados do tipo GpN (onde N pode ser qualquer um dos quatro desoxirribonucleotídeos A, C, G e T; a interação ocorre principalmente quando N é uma desoxicitidina – C). Já uma de suas porções pentapeptídicas (2 anéis, α e β , formados por treonina, D-valina, prolina, sarcosina e metil-valina) fica internamente ao sulco menor do DNA, fazendo fortes ligações de hidrogênio com os resíduos de desoxiguanosina em cadeias opostas, bloqueando a síntese de RNA associada ao DNA por impedir a atuação da enzima RNA polimerase. O autor supracitado propõe uma ligação da actinomicina ao DNA em uma conformação de pré-pareamento, chamada β -DNA (diferente da conformação comum B-DNA), encontrada no complexo de transcrição, ocasionando a inibição do processo de alongação da molécula de RNA (SOBELL, 1985; HOU *et al.*, 2002; KOBAYASHI & KONOPKA, 2005). Essa droga é utilizada costumeiramente em estudos avaliando o decaimento de mRNA em eucariotos. Para tripanossomatídeos, foi sugerida a necessidade de uma segunda droga, o inibidor de metilação sinefungina,

A meia-vida de um RNA mensageiro é o tempo necessário para que metade da sua população seja degradada. O tempo de integridade dessas moléculas determina a probabilidade de sua tradução em proteína e, conseqüentemente, a expressão de um gene. O decaimento de mRNA, que é a etapa de controle da degradação dessas moléculas, se encontra no final de um processo interno a uma rede de controles da expressão gênica. Ou seja, o decaimento representa o resultado final de um conjunto de eventos de controle pós-transcricional da expressão dos produtos gênicos e dos fenótipos que derivam desses produtos que, no caso dos mRNAs, são as proteínas (EHLERS *et al.*, 1987).

Alguns estudos avaliaram o nível de decaimento dos RNAs mensageiros, utilizando a técnica de microarranjo em bactérias (BERNSTEIN *et al.*, 2002; HAMBRAEUS *et al.*, 2003), arqueobactérias (ANDERSSON *et al.*, 2006) e eucariotos (WANG *et al.*, 2002, RAGHAVAN *et al.*, 2002; YANG *et al.*, 2003). De forma geral, esses trabalhos demonstram a importância do decaimento do mRNA como processo regulador da expressão gênica e ressaltam que o controle da taxa de degradação dos mRNAs modifica de forma rápida e eficiente a população de transcritos no citoplasma.

Recentemente, SHOCK *et al.* (2007) publicaram a primeira análise do decaimento de mRNAs de forma ampla em um protozoário, o *Plasmodium falciparum*. Interessantemente, eles avaliaram também os ritmos de decaimento do mRNA em diferentes estágios do processo de diferenciação desse parasito e identificaram mudanças grandes do mesmo na fase intra-eritrocitária assexuada.

Diferentemente dos modelos citados acima, que apresentam predominantemente regulação da expressão gênica transcricional, o *T. cruzi* regula o nível de seus mRNAs gene-específicos de forma pós-transcricional. Portanto, a determinação dos ritmos de decaimento do mRNA, de forma direta e comparativa, contém um potencial informativo ainda maior do que o visto nos demais modelos já estudados.

Até o momento, não foi realizado nenhum estudo sistemático focado no ritmo de decaimento de mRNAs em *T. cruzi*. Existem somente descrições pontuais de meia-vidas de mRNAs relativos a poucos genes para alguns tripanossomatídeos (EHLERS *et al.*, 1987, BACCHI *et al.*, 1989; ARGAMAN *et al.*, 1994; ABUIN *et al.*, 1999; VANHAMME *et al.*, 1999; WESTON *et al.*, 1999; THOMAS *et al.*, 2000; COUGHLIN *et al.*, 2000; BRITTINGHAM *et al.*, 2001; CHAUDHURI *et al.*, 2002;

GENTIL *et al.*, 2009) ou estudos sobre alterações morfológicas e bioquímicas ao se utilizarem agentes inibidores da transcrição (DA CRUZ *et al.*, 1991; MUKHOPADHYAY & MADHUBALA, 1995; ZAVERUCHA DO VALLE *et al.*, 2003; DOSSIN & SCHENKMAN, 2005; NÁZER *et al.*, 2011; NÁZER & SÁNCHEZ, 2011).

Visando preencher essa lacuna e aumentar a quantidade de informações contidas no banco de dados de transcriptomas do projeto Reguloma de *T. cruzi*, fizemos a quantificação dos níveis de mRNAs em epimastigotas tratados com actinomicina D e sinefungina pela técnica de RNA-Seq utilizando a plataforma de sequenciamento SOLiD.

3 OBJETIVOS

O presente trabalho procura aprofundar o conhecimento de como o protozoário estudado regula a expressão de seus genes, observando uma importante etapa de controle da mesma: o decaimento das moléculas transcritas codificadoras para proteínas.

3.1 OBJETIVO GERAL

- Obter o padrão de decaimento dos mRNAs de epimastigotas de *Trypanosoma cruzi* durante a fase epimastigota do seu ciclo de vida.

3.2 OBJETIVOS ESPECÍFICOS

- Sequenciar o transcriptoma de epimastigotas das cepas Dm28c e Y de *T. cruzi* tratados com actinomicina D através da plataforma SOLiD de sequenciamento em larga escala;
- Avaliar o padrão de decaimento dos RNAs mensageiros sequenciados obtidos a partir das culturas celulares;
- Identificar a equação matemática de decaimento de cada gene;
- Identificar características descritivas das diferentes categorias gênicas de decaimento.

4 MATERIAIS E MÉTODOS

4.1 CULTIVO DE PARASITOS

As culturas axênicas de *T. cruzi* clone Dm28c e cepa Y foram mantidas a 28 °C e, a cada três dias, foram repicadas mediante inóculos de 1×10^6 células por mL de solução. Para isso, epimastigotas em fase logarítmica de crescimento são cultivados em meio LIT B (*Liver Infusion Tryptose*, meio de cultura infusão de fígado e triptose) acrescido de soro fetal bovino (CAMARGO, 1964). O meio LIT B é composto por extrato de levedura 15 g/L, fosfato dibásico de sódio 11,56 g/L, glicose 2,2 g/L, hemina 0,02 g/L, infuso de fígado 5 g/L, triptose 5 g/L, KCl 0,4 g/L e NaCl 4,4 g/L. Foi adicionado soro fetal bovino 10 % (v/v) e o pH da solução deve ser ajustado para 7,2 com HCl. Assim, as formas epimastigotas em fase exponencial de crescimento foram obtidas no terceiro dia de cultivo, quando a densidade celular é de aproximadamente 2×10^7 células por mL de solução.

4.2 INIBIÇÃO DA TRANSCRIÇÃO

Para inibir a transcrição dos parasitos, a droga actinomicina D (Sigma-Aldrich; n.º de catálogo: A1410-10MG) foi diluída em dimetil-sulfóxido (DMSO 100%) em uma concentração final de $1 \mu\text{g} / \mu\text{L}$ e, então, adicionada à cultura celular visando uma concentração final de $10 \mu\text{M} / 1 \text{ mL} / 1 \times 10^7$ células. O inibidor sinefungina (Sigma-Aldrich; n.º de catálogo: S8559-10MG) foi utilizado na concentração de $1 \mu\text{g} / \text{mL}$ e colocado 5 minutos antes do ponto considerado “zero” para actinomicina D (ponto “-5 minutos”, conforme sugerido pelo protocolo de COLASANTE *et al.*, 2007). As culturas foram mantidas no meio com actinomicina e

sinefungina enquanto eram retiradas alíquotas com aproximadamente 1×10^8 células nos pontos de 0 (inclusão de sinefungina), 5 (adição de actinomicina D), 10, 15, 20 e 25 minutos. Um minuto antes do tempo nominal, a amostra era centrifugada a $13.000 \times g$ e imediatamente processada para a extração do mRNA total. Os 6 pontos temporais foram feitos com 2 réplicas biológicas, totalizando 12 amostras por cepa do parasito utilizada.

4.3 ISOLAMENTO DO MRNA TOTAL

Para extrair as moléculas de RNA de *T. cruzi*, cada amostra de ponto temporal com 1×10^8 epimastigotas foi processada com reagentes do kit *RNeasy* (Qiagen), conforme as instruções do kit, utilizando o protocolo “*Animal Cells I*” com inclusão do passo opcional de tratamento da amostra com DNase na própria coluna de extração (reagente “*RNase-free DNase set*”).

A partir das amostras de RNA total de todos os pontos temporais, utilizou-se (duas vezes consecutivas) o kit *Poly(A) PuristTM* (Applied Biosystems) para purificação da fração de RNAs poli-A⁺ das amostras. Após separação dos RNAs poli-adenilados, realizaram-se as etapas de preparo de amostras para entrada na plataforma de sequenciamento SOLiD.

4.4 PREPARO DO MATERIAL PARA A REAÇÃO DE SEQUENCIAMENTO

A partir de 500 ng de mRNA poli-A⁺ e com a qualidade checada pelo uso do equipamento *Bioanalyzer* (Agilent), procedeu-se à fragmentação por RNase III (Applied Biosystems), utilizando-se 1 μ L de RNase III e 1 μ L de tampão em um volume final de 10 μ L. A reação é incubada em um termociclador a 37 °C por 10 minutos e, após esse tempo, adiciona-se imediatamente 90 μ L de água sem

nucleases. O RNA é limpo através do reagente *RiboMinus Concentration Module* (Invitrogen) e a qualidade do RNA é analisada no *Bioanalyzer*.

Após a fragmentação dos RNAs, foram ligados adaptadores P1 e P2 aos fragmentos, disponíveis no kit *Whole Transcriptome* (Applied Biosystems), sendo 2 µL da solução de adaptadores, 3 µL da solução de hibridação e 3 µL das amostras de RNA fragmentado. O material foi colocado em um termociclador a 65 °C por 10 minutos e 16 °C por 5 minutos. Adicionou-se 10 µL de tampão de ligação 2x e 2 µL da mistura da enzima Ligase e essa solução final foi mantida a 16 °C por 16 horas.

Após a ligação dos adaptadores, foi feita a reação de transcrição reversa, com os reagentes do kit *Whole Transcriptome*. Aos 20 µL da reação de ligação, adicionou-se 13 µL de água livre de nucleases, 4 µL de tampão RT 10x, 2 µL de dNTP (desoxirribonucleotídeo) 2,5 mM, e 1 µL de transcriptase reversa. Após, as amostras foram vortexadas e incubadas em um termociclador a 42 °C por 30 minutos. Após o término da reação de transcrição reversa, purificou-se o material com o kit *MinElute PCR Purification* (Qiagen) de acordo com as instruções do fabricante.

Em seguida, fez-se a etapa de seleção do tamanho dos fragmentos de cDNA, usando o sistema *Novex* de gel com uréia e TBE (Invitrogen), com excisão da banda do gel com o tamanho adequado (de 150 a 250 pb) com lâmina de bisturi, sendo que cada fragmento de gel foi novamente recortado em quatro fatias verticais. As duas fatias, das quatro, das laterais foram congeladas e as duas outras fatias, referentes ao meio do fragmento de gel inicial, foram amplificadas por PCR, com os reagentes do kit *Whole Transcriptome*. Para cada fatia, foram adicionados 76,8 µL de água sem nucleases, 10 µL de tampão de PCR 10x, 8 µL de dNTPs, 2 µL do SOLiD 5' PCR primer, 1,2 µL de AmpliTaq DNA polimerase e 2 µL do SOLiD 3'PCR primer do kit SOLiD RNA Barcoding, sendo que para cada uma das 12 amostras de cada cepa, foram utilizadas etiquetas diferentes. As condições de PCR foram: um ciclo inicial a 95°C por 5 minutos, 15 ciclos de três etapas, sendo 95°C por 30 segundos, 62°C por 30 segundos e 72°C por 30 segundos, e uma etapa final de extensão a 72°C por 7 minutos. O material amplificado foi purificado com o kit *PureLink PCR Micro* (Invitrogen) e sua qualidade final foi verificada com o uso do *Bioanalyzer*.

4.5 REAÇÃO DE AMPLIFICAÇÃO EM EMULSÃO

O material preparado no item anterior foi amplificado em uma reação de PCR em emulsão (ePCR), na qual as moléculas de DNA são aprisionadas, juntamente com partículas magnéticas e reagentes de PCR, em micelas de óleo, formando células reatoras. As doze amostras referentes a cada cepa de *T. cruzi* foram reunidas em uma única biblioteca criando-se, portanto, duas bibliotecas diferentes que foram amplificadas separadamente. Essa PCR foi preparada utilizando-se os reagentes do kit *ePCR* (Applied Biosystems), de acordo com seu manual. Após a realização da ePCR, os reatores foram desfeitos (a emulsão foi quebrada), utilizando-se 2-butanol e as partículas magnéticas foram lavadas com o tampão exigido pelo protocolo, para eliminar o restante do óleo de emulsão. Após homogeneizar a solução, esta foi centrifugada a 2.200g por 5 minutos e então o sobrenadante foi descartado. O *pellet* contendo as partículas com os *amplicons* passou pela etapa de enriquecimento das partículas magnéticas, utilizando o kit *Bead Enrichment* (Applied Biosystems), que consiste em aumentar a concentração de partículas carregadas com material sequenciável, eliminando-se as partículas vazias.

4.6 PREPARO DAS LÂMINAS E REAÇÃO DE SEQUENCIAMENTO

Para o depósito das partículas magnéticas nas lâminas que são inseridas no equipamento de sequenciamento SOLiD (Applied Biosystems), foram utilizados os reagentes do kit *Bead Deposition* (Applied Biosystems). Primeiramente, as extremidades 3' do material aderido às partículas magnéticas foram modificadas para se tornarem capazes de se ligar à superfície da lâmina, de acordo com as recomendações do fabricante. O material (as 24 amostras temporais) foi depositado na lâmina e, esta, colocada no sequenciador. Os kits utilizados para a reação de sequenciamento foram o *Fragment Library Sequencing, Workflow Analysis* e

Fragment Library Sequencing Barcoding (todos do mesmo fornecedor do equipamento, Applied Biosystems). O processo de sequenciamento e sua qualidade foram avaliados utilizando-se os *softwares* SAET e SETS (Applied Biosystems).

4.7 MAPEAMENTO DAS SEQUÊNCIAS

Primeiramente, os resultados do sequenciamento foram avaliados utilizando-se o *software* SOLiD™ *WT Analysis Pipeline* (Applied Biosystems) e as sequências com 50 nucleotídeos foram mapeadas, com o mesmo *software*, no genoma referência do *T. cruzi* (cepa CL Brener). Previamente ao mapeamento das leituras, o próprio *software* da máquina elimina sequências consideradas de baixíssima qualidade.

Para o mapeamento é necessária a definição do limite dos genes, que no caso do genoma do *T. cruzi*, não é sabido. Portanto, foi necessário definir o que é considerado como “gene” *versus* “região intergênica”. Somente os *reads* que são mapeados em regiões internas a algum “gene” são computados nas etapas posteriores. O fluxograma com todas as opções de processamento para a realização do mapeamento que foram testadas está representa na FIGURA 4.1 abaixo.

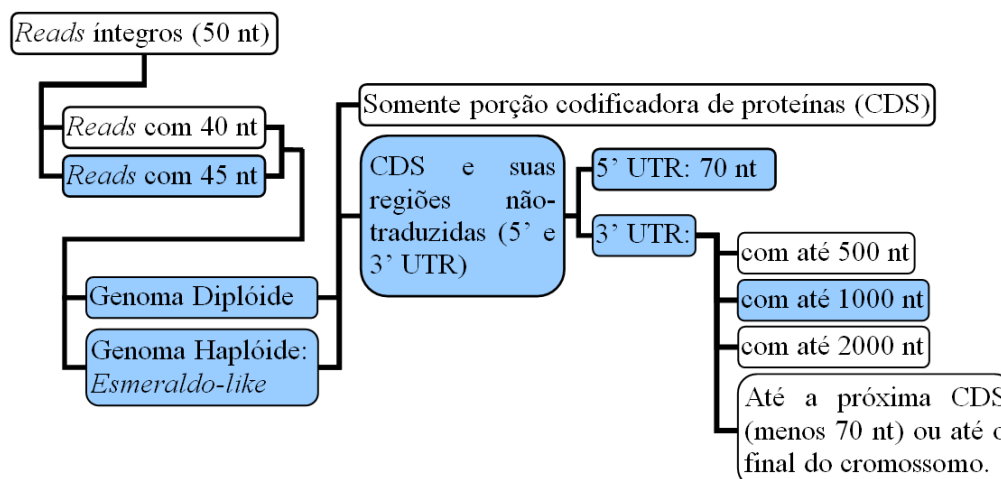


FIGURA 4.1. ORGANOGrama DOS TIPOS DE MAPEAMENTOS TESTADOS.

Em destaque a forma de mapeamento considerada otimizada para os dados obtidos.

FONTE: O autor (2011).

Nenhum gene foi considerado como possuindo limites que sobrepunham outros genes. No máximo, a proximidade do início de um gene se dá no nucleotídeo seguinte após o nucleotídeo que delimita o término do gene que se encontra imediatamente a jusante do primeiro. De todas as formas possíveis mostradas na Figura 4.1, após comparações entre o resultado do número de *reads* mapeadas de acordo com essas definições de limites dos genes, apenas 4 foram selecionadas para seguirem os procedimentos de análise, que são os tipos de mapeamento:

- **Genoma diplóide, UTR1000 (sigla: “UTR”)**: considera ambos haplótipos da cepa CL Brener, e a região gênica é composta pela CDS e pelas sequências até 70 nt a 5’ da CDS (menor quando existe sobreposição com a CDS adjacente a 5’) e até 1000 nt a 3’, desde que não ultrapasse o início da CDS seguinte (a jusante) e sua 5’ UTR de 70 nt.
- **Genoma diplóide, CDS (sigla: “CDS”)**: considera somente a porção codificadora (CDS), sem as UTRs, de todos os genes de ambos haplótipos da cepa CL Brener.
- **Genoma Esmeraldo-like, UTR1000 (sigla: “UTRe”)**: considera somente o haplótipo Esmeraldo-like da cepa CL Brener, e a região gênica é composta pela CDS, com sua região 5’ UTR de até 70 nt a montante e sua região 3’ UTR de até 1000 nt a jusante da CDS.
- **Genoma Esmeraldo-like, CDS (sigla: “CDSe”)**: considera somente o haplótipo Esmeraldo-like da cepa CL Brener, e a somente a CDS, sem adicionar nenhuma região UTR possível.

Um detalhe sobre o mapeamento é que só são computadas as *reads* que mapearem em até 10 regiões do genoma, as demais são descartadas. Um *script* em linguagem Perl foi criado para separar em arquivos diferentes os genes codificadores de proteínas (que geram mRNAs) dos genes para ncRNAs (RNAs não codificadores), como os RNAs pequenos nucleares (snRNAs), RNAs pequenos nucleolares (snoRNAs), RNAs ribossômicos (rRNAs), entre outros. Foi escrito esse algoritmo para selecionar apenas os genes codificadores de proteínas pelo fato de que foi purificada a fração poli-A⁺ dos RNAs e o foco do trabalho se dá aos mRNAs. Esse *script* utiliza palavras-chave para distinguir os genes nessas categorias segundo sua anotação. A eficiência do *script* foi testada manualmente, verificando quais genes eram separados como ncRNAs.

4.8 NORMALIZAÇÃO E FILTRAGENS DOS DADOS

A normalização das contagens de sequências mapeadas para cada gene foi realizada pelo método TMM (média ajustada, diminuída, dos valores de M, do inglês “*trimmed mean of M values*”; ROBINSON & OSHLACK, 2010) que, basicamente, é um método de normalização aplicável entre bibliotecas (réplicas das amostras temporais em separado) que ajusta as mesmas em consideração o seu tamanho (número total de sequências mapeadas) e leva em consideração a intensidade da variação entre as réplicas dos valores de M (log. do “*fold-change*”). Para esse tipo de normalização entre os tamanhos das bibliotecas, e conseqüentemente entre a diferença de composição das mesmas, usamos o pacote *edgeR* (ROBINSON *et al.*, 2010, o pacote está disponível no site: “<http://www.bioconductor.org/help/bioc-views/release/bioc/html/edgeR.html>”), do ambiente de análises estatísticas e linguagem de programação R (“www.r-project.org”). Este e outros pacotes (*geneplotter*, *bioDist*, *limma* de SMITH, 2004, *annotate*) da mesma iniciativa (*Bioconductor*, DUDOIT *et al.*, 2003 e GENTLEMAN *et al.*, 2004, disponibilizados em “www.bioconductor.org”) foram utilizados dentro da plataforma de análises R para gerar gráficos e matrizes, e para o tratamento e as análises dos dados. Adicionalmente, criamos alguns códigos-fonte na linguagem R e diversos *scripts* em linguagem de programação Perl para manipulações mais específicas dos dados.

Obtém-se como resultado final a mesma tabela de contagem de sequências que mapeiam a uma determinada região do genoma desse organismo, porém, após a normalização, seus valores deixam de ser discretos e passam a ser contínuos.

Feita a normalização, dois filtros de ruído dos dados foram utilizados. O primeiro, denominado linha de base, ajustou para cinco todas as contagens inferiores a esse valor, para diminuir a flutuação da expressão de genes pouco representados. Já que estes são os genes mais suscetíveis ao ruído de fundo inerente das técnicas utilizadas e dificultando a captura do sinal dos fenômenos biológicos passíveis de detecção. O segundo, denominado filtro de médias retirou das análises posteriores todos os genes com média de contagens menor que 15.

Assim foram definidas as quantidades de transcritos encontrados para cada gene anotado no genoma, que representam uma medida quantitativa da expressão dos genes.

4.9 IDENTIFICAÇÃO DO PADRÃO DE DECAIMENTO

Os genes expressos foram classificados em grupos conforme a meia-vida de seus transcritos, dos mRNAs mais estáveis aos de rápido decaimento.

O decaimento gene específico foi modelado matematicamente no R por regressão linear das contagens normalizadas e filtradas, gerando a função de 1º grau que cria a reta que melhor explica os pontos num gráfico com as contagens em função da série temporal, conforme a EQUAÇÃO 4.1 a seguir.

$$Y = f(X) = \alpha + \beta X \quad ; \quad y = f(x) = a + bx + \varepsilon$$

EQUAÇÃO 4.1. EQUAÇÃO DA RETA DE REGRESSÃO LINEAR “F(X)”.

À esquerda a notação populacional e à direita a amostral.

Onde Y é o número de transcritos de um dado gene, que está em função do tempo X . Considerando que estamos tomando medidas das variáveis reais, por um estudo amostral, y é o número de sequências, x é o tempo medido e ε é o erro da predição. A variável a determina o ponto em que a reta de regressão linear cruza o eixo das ordenadas e a variável b a inclinação dessa reta de regressão. A fim de avaliar a adequação da fórmula aos dados reais, foi calculado o coeficiente de variação dos resíduos das regressões lineares, conforme a EQUAÇÃO 4.2:

$$\text{C.V.R.}_{\text{gene}} = \frac{\begin{array}{c} \text{Desvio padrão dos resíduos} \\ \text{em todas as amostras} \\ \text{média das contagens} \\ \text{em todas as amostras} \end{array}}{\bar{y}_{\text{gene}}} = \frac{\sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-2}}}{\bar{y}_{\text{gene}}}$$

EQUAÇÃO 4.2. COEFICIENTE DE VARIAÇÃO DOS RESÍDUOS (CVR).

Na qual y_i é a contagem para a amostra i , \hat{y}_i é a regressão para o ponto y_i , ou seja, a projeção do ponto y_i na reta. Os graus de liberdade dos resíduos é determinado pelo número de amostras (n) menos duas unidades, ou seja, igual a 10.

Para comparação da reprodutibilidade entre as duas réplicas técnicas, usamos o coeficiente de correlação de Pearson (" r "), conforme a EQUAÇÃO 4.3.

$$r = \frac{1}{n-1} \sum_{i=1}^n \left[\left(\frac{X_i - \bar{X}}{\sigma_X} \right) \left(\frac{Y_i - \bar{Y}}{\sigma_Y} \right) \right]$$

EQUAÇÃO 4.3. COEFICIENTE DE CORRELAÇÃO DE PEARSON (" R ").
(ou coeficiente de correlação produto-momento, centrado por e dependente da média).

Essa medida de correlação varia entre correlação negativa total (-1) e correlação positiva total (1), passando pela ausência de correlação entre as variáveis (0).

A inclinação da reta (valor de b da regressão linear) foi utilizada para outras análises. Os valores da inclinação das retas de regressão linear dos genes foram ordenados e as extremidades dessa distribuição foram divididas em 10 grupos (ou *bins*) de genes, de acordo com dois procedimentos: *bins* contendo 100 ou 500 genes cada.

4.10 CLUSTERIZAÇÃO

As contagens dos genes na série temporal foram organizadas por clusterização hierárquica aglomerativa (método UPGMA, do inglês *Unweighted Pair Group Method with Arithmetic Mean*, de SOKAL & MICHENER, 1958) e por clusterização *k-means* (com $k = 8$), utilizando a mesma métrica de distância do método de Eisen (que calcula a correlação, não-centrada pela média, entre os pontos da série temporal, vide EQUAÇÃO 4.4) e seu respectivo *software* "*Cluster*"

(EISEN *et al.*, 1998). Com os *reads* mapeados e anotados puderam ser formuladas e testadas hipóteses sobre as implicações fisiológicas do decaimento.

$$S(X,Y) = \frac{1}{N} \sum_{i=1}^N \left(\frac{X_i - X_{offset}}{\Phi_X} \right) \left(\frac{Y_i - Y_{offset}}{\Phi_Y} \right)$$

$$\Phi_G = \sqrt{\sum_{i=1}^N \frac{(G_i - G_{offset})^2}{N}}$$

EQUAÇÃO 4.4. MÉTRICA DE SIMILARIDADE DO MÉTODO DE CLUSTERIZAÇÃO DE EISEN ET AL. (1998).

A métrica de similaridade (ou métrica de distância) de Eisen se comporta como uma medida de correlação de Pearson.

Para visualização dos *clusters* formados e confecção das respectivas imagens mostradas nesse trabalho, utilizamos o *software* “Java TreeView” (disponível em “<http://jtreeview.sourceforge.net>”)

4.11 CATEGORIZAÇÃO FUNCIONAL

Para identificar os grupos funcionais mais característicos relativos às diferentes análises realizadas, utilizamos a ferramenta *Blast2GO* (CONESA *et al.*, 2005). Resumidamente, a ferramenta foi utilizada para atribuir termos relativos à iniciativa *Gene Ontology* (GO) aos genes de *T. cruzi*, através da identificação de ortólogos por BLAST e de domínios funcionais, pelo InterProScan. Após a atribuição dos termos GO aos genes, foram realizados análises de enriquecimento funcional pelo teste exato de Fisher com correção para múltiplos testes por FDR.

5 RESULTADOS E DISCUSSÃO

Uma das principais características de um trabalho de genômica funcional é a obtenção de grande quantidade de dados, os quais devem ser analisados sob diferentes perspectivas para obtermos uma compreensão melhor sobre o assunto estudado. Tratando-se de padrões de decaimento de mRNA, é necessário avaliar o conjunto como um todo, sendo exigido muita cautela em relação à influência do ruído, entre outros artefatos, que podem distorcer os dados de maneira pouco perceptível. A fim de tornar o entendimento das análises feitas em relação ao conjunto dos dados mais acessíveis, optamos por proceder a apresentação dos resultados associada à discussão dos mesmos.

5.1 QUALIDADE DAS AMOSTRAS DE MRNA

A análise da qualidade do RNA total obtido das amostras coletadas foi processado no equipamento *2100 Bioanalyser* (Agilent) (FIGURA 5.1), na qual se identifica vários picos no gráfico. Os três picos maiores são referentes aos rRNAs 24Sa, 24Sb e 18S, sendo o perfil eletroforético característico de tripanossomatídeos. De maneira geral, o padrão observado é de amostras íntegras, condição essencial para a realização de um trabalho em larga escala; todas as amostras de RNA total analisadas no presente trabalho possuem o mesmo perfil.

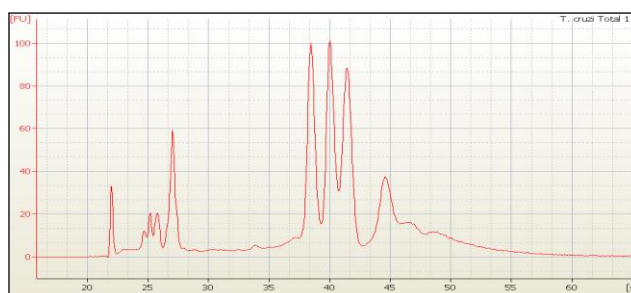


FIGURA 5.1. PERFIL ELETROFORÉTICO COMUM ÀS AMOSTRAS DE RNA TOTAL.
FONTE: O autor (2011).

5.2 QUALIDADE DO SEQUENCIAMENTO

Todas as 24 amostras do experimento de tratamento dos parasitos com actinomicina D e sinefungina foram preparadas conforme os protocolos citados anteriormente e sequenciadas pelo sequenciador SOLiD. Os resultados do sequenciamento foram avaliados com relação à qualidade das sequências obtidas e notamos que tanto o sequenciamento das amostras da cepa Dm28c quanto das amostras da cepa Y apresentavam qualidade bem acima do limite mínimo exigido (segundo as configurações automáticas do *software* SAET) e com suficiente heterogeneidade de sequências (conforme o *Satay Plot* do *software* SETS).

5.3 QUALIDADE DAS SEQUÊNCIAS

Apesar das amostras passarem nesses critérios de qualidade exigidos durante a etapa do sequenciamento, foi observado que as sequências obtidas de Dm28c apresentavam problemas. Dentre eles, há evidências da não ocorrência de decaimento, maior irreprodutibilidade entre as réplicas e mudança do perfil de decaimento ao longo do processo.

Esperava-se que, na medida em que a cultura celular fosse exposta às drogas, a distribuição dos transcritos referente à sua taxa de decaimento entre os diferentes pontos temporais tenderia a se distanciar das medidas de tendência central (média, moda e mediana). Nos dados de Y isso é observado (FIGURA 5.2), porém os de Dm28c são divergentes aos de Y, não apresentando o padrão de aumento gradual de decaimento, sendo que para a comparação entre os pontos 10 e 20 o padrão é semelhante a uma distribuição aleatória; e entre 5 e 15 minutos o padrão é muito semelhante a uma distribuição normal, não apresentando a característica bimodalidade vista em Y (FIGURA 5.3).

Nas distribuições dos valores das regressões lineares dos genes nos dados da cepa Y, percebe-se que a moda e a mediana das distribuições se deslocam para

a esquerda (maioria dos transcritos em decaimento ao passar do tempo) e a densidade de genes com valor de “ R^2 ” (valor da regressão linear) próximo a zero diminui. Essas características das distribuições dos valores das regressões dos genes fortalecem um perfil de decaimento de transcritos em amostras normalizadas pela quantidade de RNA.

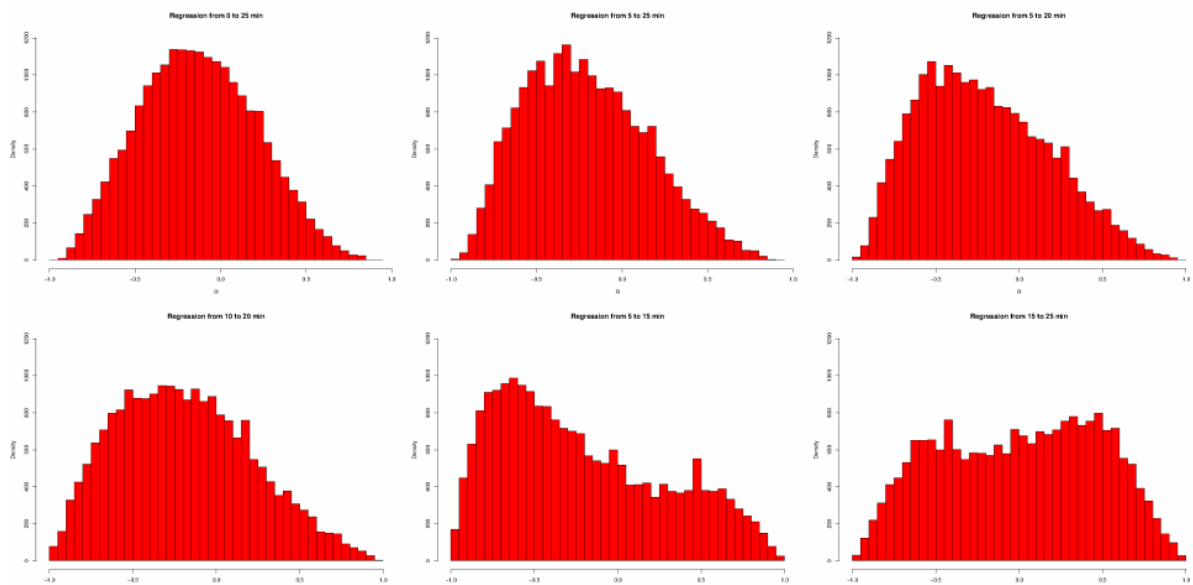


FIGURA 5.2. HISTOGRAMA DOS VALORES DE “ R^2 ” PARA AS AMOSTRAS DA CEPA Y.

O valor de “ R^2 ” (eixo “x”) da regressão linear das contagens de sequências reflete o decaimento de cada gene durante a série temporal. A série toda (de 0 minuto até 25 minutos) ou trechos da mesma estão representados. Da esquerda para a direita, acima, são considerados os pontos: de 0’ a 25’, de 5’ a 25’ e de 5’ a 20’. Abaixo, de 10’ a 20’, de 5’ até 15’ e de 15’ a 25’. No eixo “y” está a densidade de genes para cada intervalo. FONTE: O autor (2011).

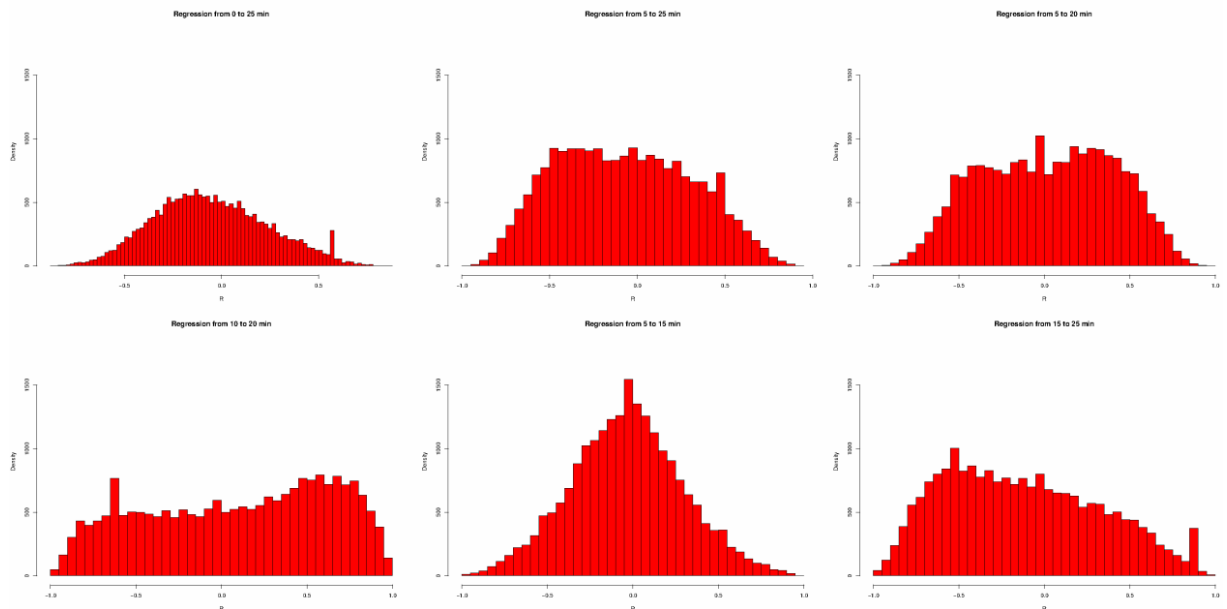


FIGURA 5.3. HISTOGRAMA DOS VALORES DE “ R^2 ” PARA AS AMOSTRAS DA CEPA DM28C.

Nota explicativa da figura idêntica à FIGURA 5.2. FONTE: O autor (2011).

Comparando as características gerais citadas entre as amostras das duas cepas, foi decidido que os dados de Dm28c não seriam analisados na presente dissertação. A tentativa de identificar a causa das diferenças entre os dados de Dm28c e Y demanda novos experimentos laboratoriais e a exploração ampla dos dados já obtidos, o que é inviável em relação ao tempo exigido para a conclusão da dissertação. Sendo assim, a partir deste ponto o presente trabalho foi realizado apenas com os dados dos experimentos com a cepa Y, contendo duas réplicas biológicas de seis amostras de RNA mensageiro dos pontos temporais já citados. O número de sequências obtidas para cada uma das doze bibliotecas se encontra na TABELA 5.1 abaixo.

Réplicas A	N.º de Reads	Réplicas B	N.º de Reads
0 minutos	21.912.297	0 minutos	14.085.835
5 minutos	12.588.021	5 minutos	16.000.583
10 minutos	16.174.480	10 minutos	13.282.207
15 minutos	26.161.153	15 minutos	14.756.135
20 minutos	25.113.693	20 minutos	14.246.378
25 minutos	15.360.694	25 minutos	19.675.823
Total A	117.310.338	Total B	92.046.961
Total A+B		209.357.299	

TABELA 5.1. CONTAGEM DE SEQUÊNCIAS OBTIDAS PARA CADA BIBLIOTECA.
 FONTE: O autor (2011).

As 209 milhões de sequências foram postas a mapear no genoma de *T. cruzi*, mas apenas cerca de 30% das sequências mapearam no genoma de CL Brener. Esse número é relativamente baixo, embora represente um total de 60 milhões de sequências mapeadas; em outros experimentos conduzidos por nosso grupo com *RNA-Seq*, a porcentagem de mapeamento alcança até 80%.

5.4 ANÁLISE DE FATORES INFLUENCIANDO A MAPEABILIDADE

Algumas razões possíveis para o não-mapeamento das sequências no genoma são (i) a complexidade do genoma de *T. cruzi* devido ao grande número de

famílias multigênicas e menor complexidade das UTRs dos genes, (ii) a divergência entre a cepa CL Brener (do genoma referência) e a cepa Y, (iii) a estringência dos critérios de mapeamento das sequências, (iv) a contaminação das amostras por mRNAs mitocondriais derivados do kDNA e editados por poli-uridinação ou por mRNAs de bactérias que contaminaram a cultura e, ainda, (v) a baixa qualidade das sequências não mapeadas e (vi) os erros de sequenciamento não capturados pelo algoritmo de determinação numérica de qualidade do *software* de análise do sequenciador.

A qualidade do sequenciamento das amostras de Y se mostrou um pouco mais alta para as sequências mapeadas do que para as não mapeadas, como evidenciado pela FIGURA 5.4, na qual é representada a distribuição da proporção de bases cujo valor de qualidade é maior do que 15 (3% de erro por nucleotídeo) em cada sequência. Embora essa diferença não seja muito forte, é significativa, pois o tamanho das sequências geradas é pequena (50 nt) e alguns poucos nucleotídeos mal sequenciados podem impedir o mapeamento das amostras. No entanto, é pouco provável que somente esse fator explique a proporção de 70% de sequências não mapeadas.

Ao repetirmos o processo de mapeamento para as sequências não mapeadas, desta vez contra o genoma mitocondrial da cepa Esmeraldo de *T. cruzi*, que não foi analisada inicialmente, conseguimos mapear somente 1,5%. Portanto, existem mRNAs mitocondriais no conjunto sequenciado, o que era esperado, no entanto essa proporção não explica a menor mapeabilidade dos dados da presente dissertação. Porém, como os RNAs derivados desse genoma são editados, essa hipótese ainda pôde ser válida. Esse processo de edição poderia ser relevante para gerar RNA não facilmente mapeáveis no genoma, especialmente se os mRNAs mitocondriais da cepa Y forem fortemente editados.

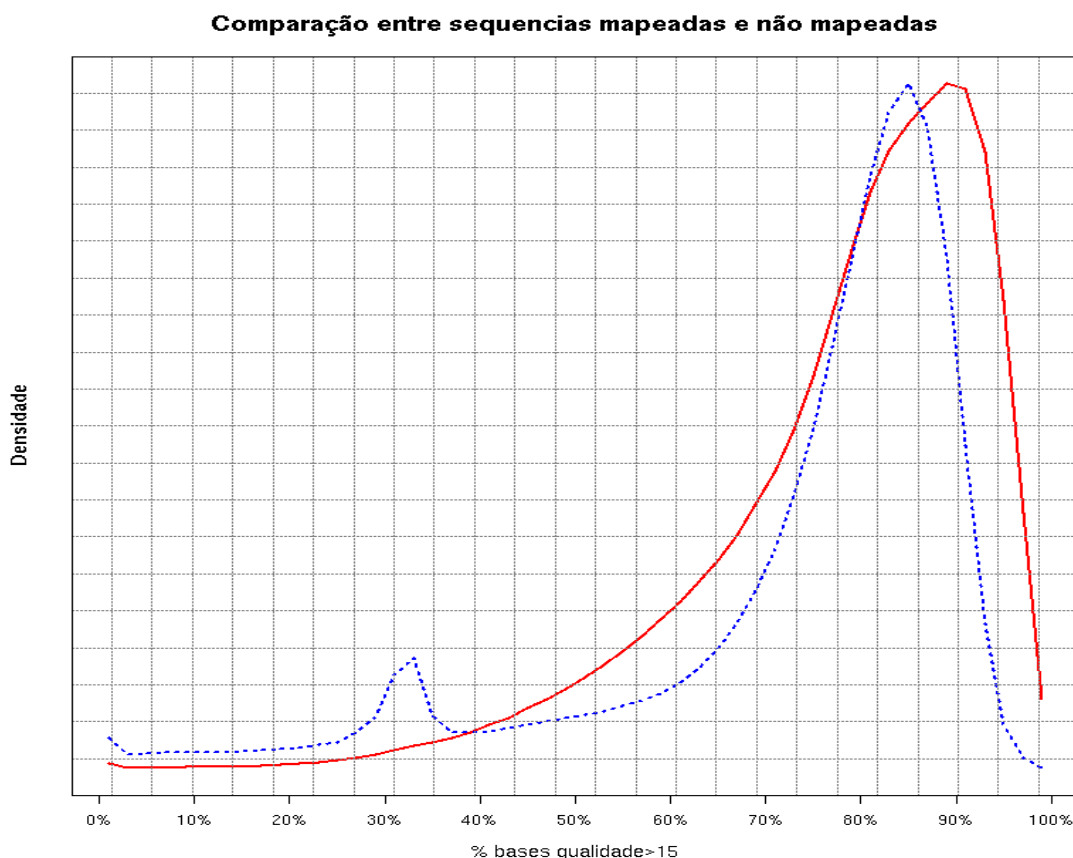


FIGURA 5.4. DENSIDADE DE SEQUÊNCIAS CUJA PORCENTAGEM DE BASES APRESENTA QUALIDADE MAIOR QUE 15.

Em vermelho (linha contínua) as sequências mapeadas e em azul (linha tracejada) as que não foram mapeadas no genoma de *T. cruzi*.

FONTE: O autor (2011).

A fim de comparar as sequências não mapeadas contra o GenBank, visando identificar possíveis contaminantes, selecionamos as sequências em que todos os primeiros 45 nt tinham valor de qualidade maior do que 15. De um total de aproximadamente 22 milhões de sequências, referentes à réplica A do ponto zero minuto, somente 131.428 passaram esse critério (0,6% do conjunto total).

Na comparação contra o banco de dados NT de sequências nucleotídicas do GenBank, 79,2% das sequências analisadas não apresentam similaridade mínima (e-value limite de 10); a maioria das sequências restantes que apresentaram similaridade é espúria; o pequeno restante apresenta similaridade com o mini-éxon, genes de famílias multigênicas, elementos retrotransponíveis e genes mitocondriais. Na comparação de uma parcela dos dados não mapeados com o banco de dados NR de sequências protéicas do GenBank, 95,1% não apresentaram similaridade

mínima (e-value limite de 5) e a maioria das similaridades identificadas nas demais sequências é espúria.

De maneira geral, a presença de transcritos mitocondriais ou a contaminação com mRNA de outros organismos podem ser excluídas como responsáveis pelo menor mapeamento observado no presente estudo. A qualidade das sequências parece ter uma importância considerável, embora não explique totalmente os problemas. A última das hipóteses listadas que permanece é a má qualidade das sequências sem haver a captura dessa propriedade no valor de qualidade do sequenciamento. É sabido que a determinação desses parâmetros de qualidade, nas novas técnicas de sequenciamento, é empírica, podendo haver falhas no algoritmo de determinação. No entanto, não temos como testar essa hipótese e ela permanece em aberto.

5.5 EFICIÊNCIA DO PROCESSO DE PURIFICAÇÃO DE MRNA

Nesse trabalho, focamos apenas nos genes codificadores de mRNAs, e retiramos os RNAs não-codificadores de proteínas, sendo que, no total, 14,07%, 9,12%, 5,52% e 0,34% das leituras mapearam em genes para ncRNAs nos mapeamentos do tipo “CDS”, “UTR”, “CDSe” e “UTRe”, respectivamente. Esses valores são demonstrativos de pouca contaminação de ncRNAs nas amostras sequenciadas, evidenciando a eficiência da purificação dupla do RNA poli-A⁺.

5.6 QUANTIFICAÇÃO DE ELEMENTOS COM BAIXA CONTAGEM

Após normalização, os dados foram processados de acordo com as estratégias descritas anteriormente (linha de base = 5 e filtro para média < 15) para diminuir o ruído nos dados e priorizar o sinal referente aos fenômenos biológicos a

serem observados. Na TABELA 5.2, encontra-se o número de células modificadas com o estabelecimento da “linha de base” e das que se mantiveram com o dado original (por possuírem mais de cinco sequências). É possível evidenciar que entre 16% e 35% das contagens era de baixo valor e que essa proporção está diretamente ligada à “mapeabilidade” dos quatro diferentes tipos de definição de regiões mapeáveis, isto é, quanto maior o número de regiões mapeáveis, menor o número de contagens modificadas. A principal explicação para essa proporção de genes com contagem baixa é o nível de expressão dos mRNAs, associado à capacidade de serem mapeados especificamente; regiões que se repetem no genoma em mais de 10 lugares são excluídos do mapeamento e isso afeta principalmente famílias multi-gênicas, que geralmente apresentam poucas sequências únicas mapeadas.

Células	CDS	%	UTR	%	CDSe	%	UTRe	%
Originais	200.257	78,97%	212.338	83,73%	84.923	68,58%	80.687	65,16%
Alteradas	53.327	21,03%	41.246	16,27%	38.905	31,42%	43.141	34,84%
Total	253.584	100%	253.584	100%	123.828	100%	123.828	100%

TABELA 5.2. NÚMERO E PORCENTAGEM DAS CONTAGENS MODIFICADAS APÓS A “LINHA DE BASE”.

FONTE: O autor (2011).

Na FIGURA 5.5 é possível observar, para os quatro tipos de mapeamento, o número de genes com média de contagem maior ou igual a diversos valores (de 1, 2, 5 e 10 até 200) dos dados normalizados antes da determinação da linha de base e do filtro de médias. Já na FIGURA 5.6, o mesmo dado depois de realizada a linha de base e a filtragem (exclusão) de genes com média menor que 15.

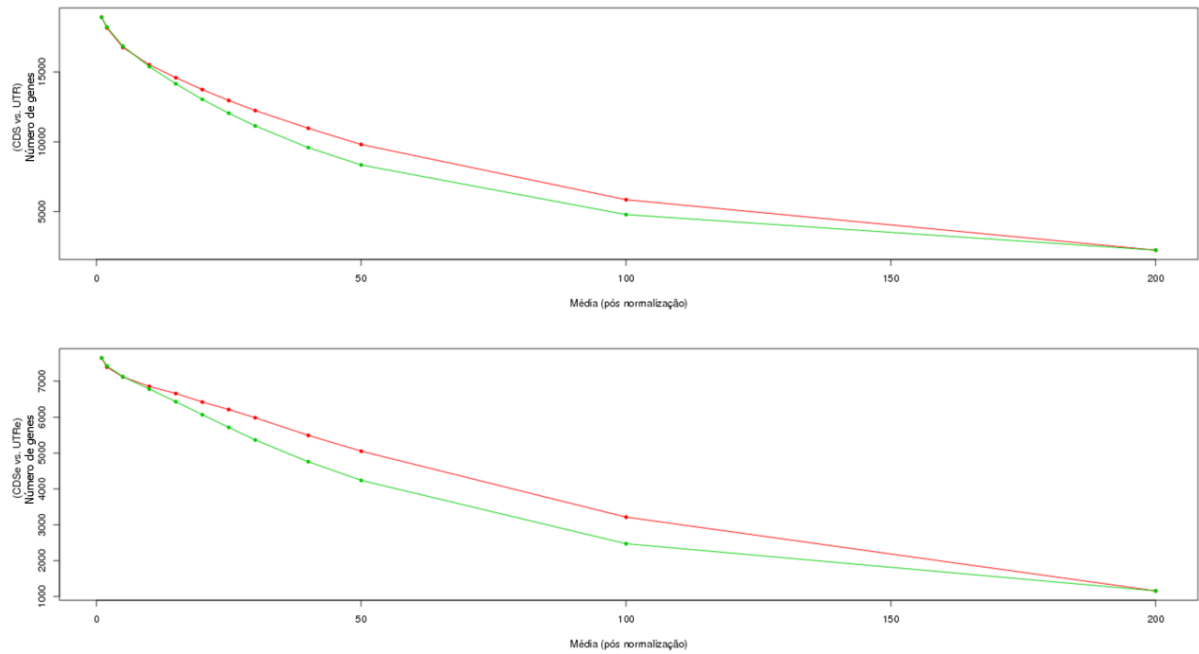


FIGURA 5.5. NÚMERO DE GENES DE ACORDO COM SUAS MÉDIAS DAS CONTAGENS DE SEQUÊNCIAS MAPEADAS.

Dados após a normalização por tamanho de amostra para os mapeamentos diplóides (CDS em verde e UTR em vermelho) acima e haplóides abaixo (mesmas cores).

FONTE: O autor (2011).

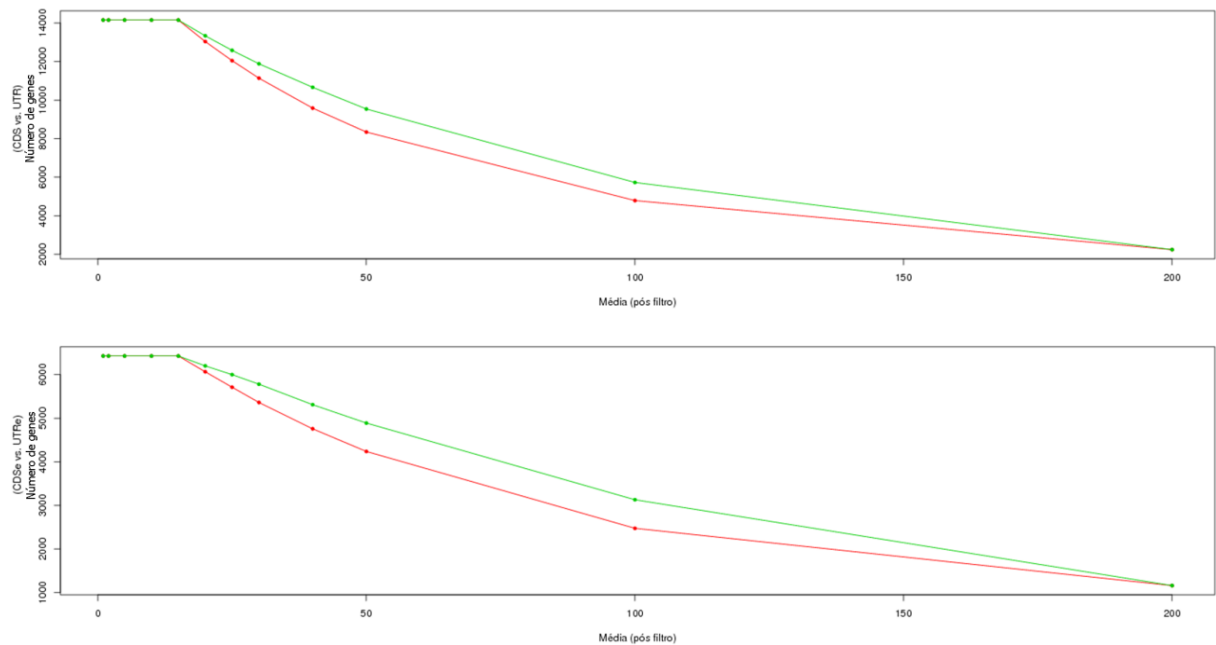


FIGURA 5.6. NÚMERO DE GENES DE ACORDO COM SUAS MÉDIAS DAS CONTAGENS DE SEQUÊNCIAS MAPEADAS

Dados posteriores ao “filtro de médias” para os mapeamentos diplóides (CDS em vermelho e UTR em verde) acima e haplóides abaixo (“CDS_e” em vermelho e “UTR_e” em verde).

FONTE: O autor (2011).

5.7 ANÁLISE DA REPRODUTIBILIDADE BIOLÓGICA

As réplicas biológicas apresentaram uma alta correlação de Pearson positiva que varia entre os pontos temporais, sendo a média para os diferentes mapeamentos testados em torno de 0,96 conforme nota-se no gráfico (mapeamento “UTR”, diplóide) com as contagens já normalizadas e filtradas, em escala logarítmica na base 2, que compara respectivamente as réplicas biológicas dos pontos temporais de 0 a 25 minutos (FIGURA 5.7).

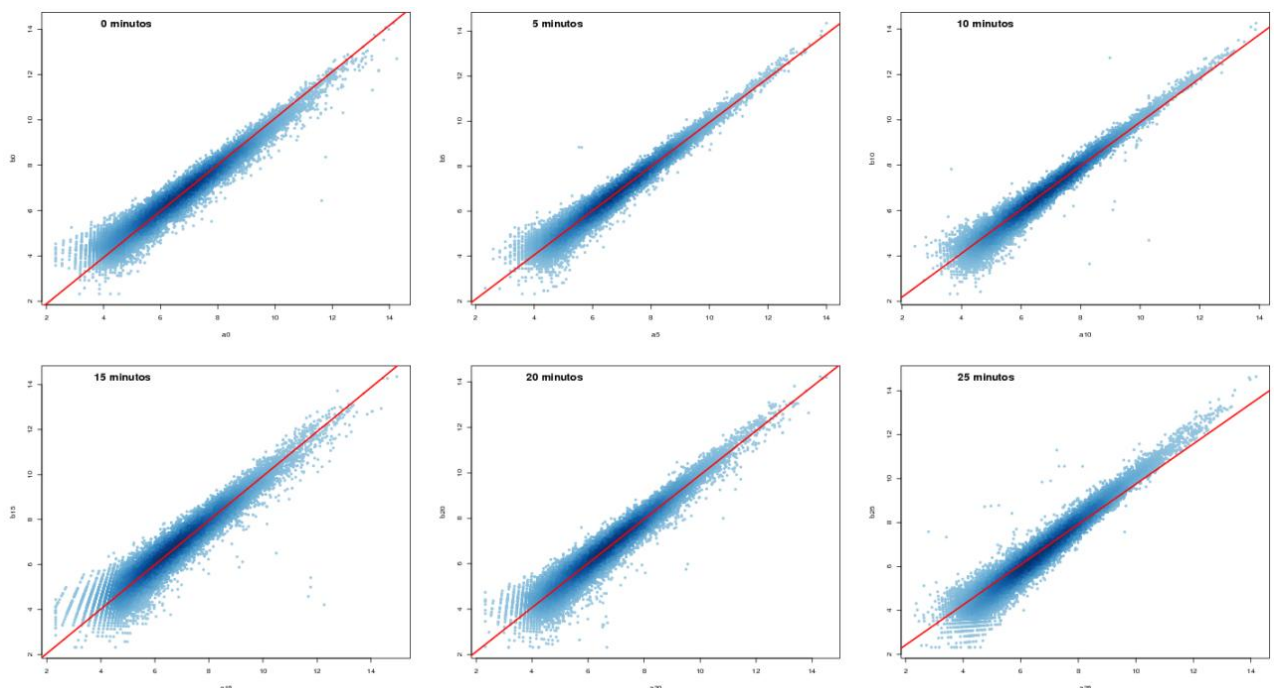


FIGURA 5.7. COMPARAÇÃO ENTRE AS CONTAGENS NORMALIZADAS E FILTRADAS DAS RÉPLICAS BIOLÓGICAS.

Um gráfico para cada amostra temporal, de 0 a 25 minutos a cada 5 minutos, do mapeamento no genoma diplóide do tipo UTR. Cada eixo representa uma réplica do ponto temporal. A linha vermelha é a reta de regressão linear.

FONTE: O autor (2011).

Uma outra forma de representar a homogeneidade e verificar as diferenças entre as amostras é com um gráfico do tipo “M versus A”, no qual é possível ver de forma mais resolutive a reprodutibilidade entre as amostras. O mesmo se trata da intensidade de sinal transformada para logaritmo na base 2 (número de contagens, ou concentração; “A”, no eixo “x”) versus o logaritmo na base 2 da razão de

mudança (*fold-change*, que é diferença da concentração de uma réplica em relação outra; “M”, no eixo “y”). Nesse tipo de gráfico podemos verificar onde se encontra a maioria dos pontos referentes ao número de contagens para cada um dos genes nas diferentes amostras da série temporal e verificamos quais amostras sofrerão maior modificação pelo fator de normalização do método TMM. Esse fator de normalização representa o ajuste aplicado às bibliotecas com relação ao seu tamanho e à diferença composicional da concentração de cada gene. Na FIGURA 5.8, observa-se que as amostras temporais de 15 a 25 são as mais divergentes (pré-normalização) entre réplicas e em relação a suas retas de regressão linear.

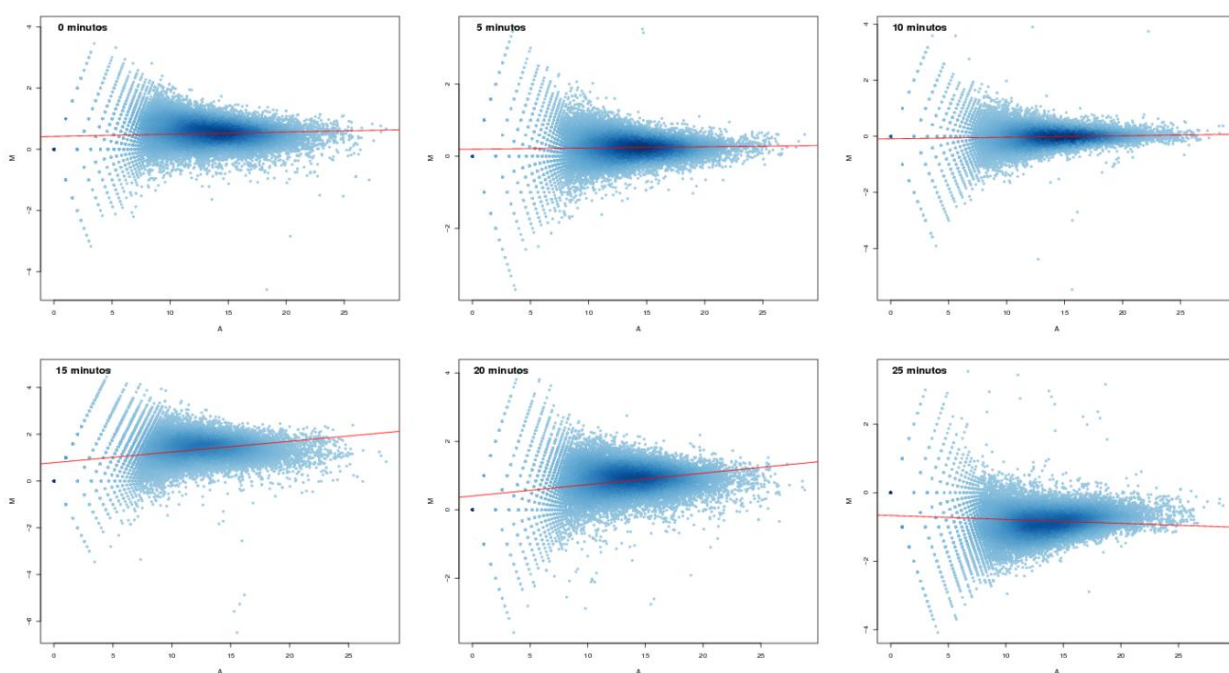


FIGURA 5.8. “MA PLOT” COMPARANDO OS DADOS BRUTOS ENTRE AS RÉPLICAS PARA CADA AMOSTRA TEMPORAL.

Um gráfico para cada par de amostra temporal, de 0 a 25 minutos a cada 5 minutos, do mapeamento no genoma diplóide do tipo UTR. No eixo “x” tem-se o valor de “A” (log. na base 2 do número de *reads*). No eixo “y” está o valor de “M” (log. de base 2 da razão de diferença ou “*fold change*”). A linha vermelha é a reta de regressão linear.

FONTE: O autor (2011).

Essas amostras foram as que sofreram maior alteração após o ajuste de normalização (FIGURA 5.9). Também é possível notar, comparando as figuras, que as três últimas amostras (principalmente as de 15 minutos) da série temporal possuem a maior densidade de genes localizados no ponto zero para ambos os eixos. Essa densidade se altera após os filtros para genes com baixa contagem, no

qual estes são retirados. Após a normalização, fica evidente que o gráfico das réplicas do tempo zero, por exemplo, se modifica um pouco. Nesse caso, isso ocorre devido à diferença do número de moléculas sequenciadas entre as duas bibliotecas do tempo zero minuto. Feita a normalização, todos os dados se aproximam do ponto zero no eixo com os valores de M (pois as diferenças da contagem dos genes entre as réplicas são ajustadas) e acabam por se concentrarem em torno da linha de regressão.

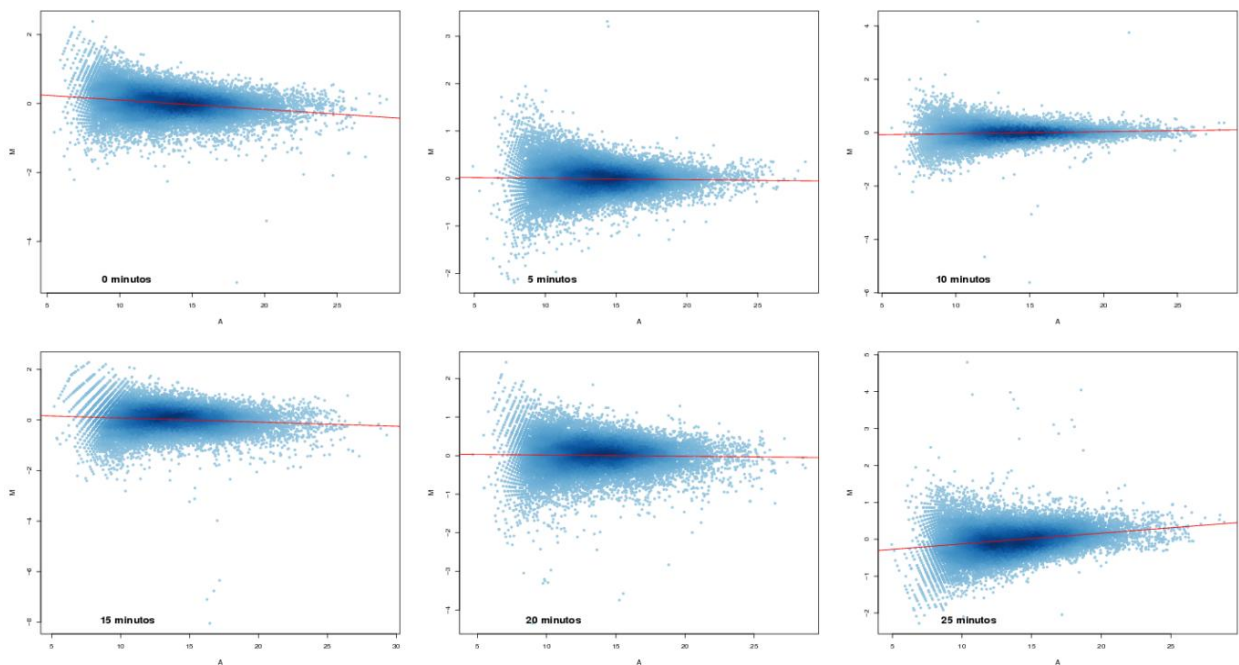


FIGURA 5.9. “MA PLOT” COMPARANDO OS DADOS NORMALIZADOS E FILTRADOS.

Um gráfico para cada par de réplicas biológicas da amostra temporal, de 0 a 25 minutos a cada 5 minutos. Mapeamento no genoma diplóide “UTR”. Os eixos são os mesmos da figura anterior. Os dados se deslocam por haver pontos distantes da linha vermelha, que é a reta de regressão linear. FONTE: O autor (2011).

5.8 ANÁLISE COMPARATIVA DOS MÉTODOS DE DEFINIÇÃO DA REGIÃO MAPEÁVEL

A fim de avaliar a influência dos métodos de mapeamento na determinação do grau de decaimento dos genes, fizemos a avaliação do valor de decaimento (*slope* ou coeficiente angular: parâmetro “b” da equação geral da reta) obtido para

cada um dos quatro métodos entre si. Caso houvesse influência da forma de mapeamento na determinação do valor de decaimento, a correlação entre pelo menos um dos métodos em relação aos outros seria muito baixa. De maneira geral, há uma forte correlação positiva dos valores de decaimento entre os diferentes métodos de mapeamento, conforme pode ser visto na FIGURA 5.10.

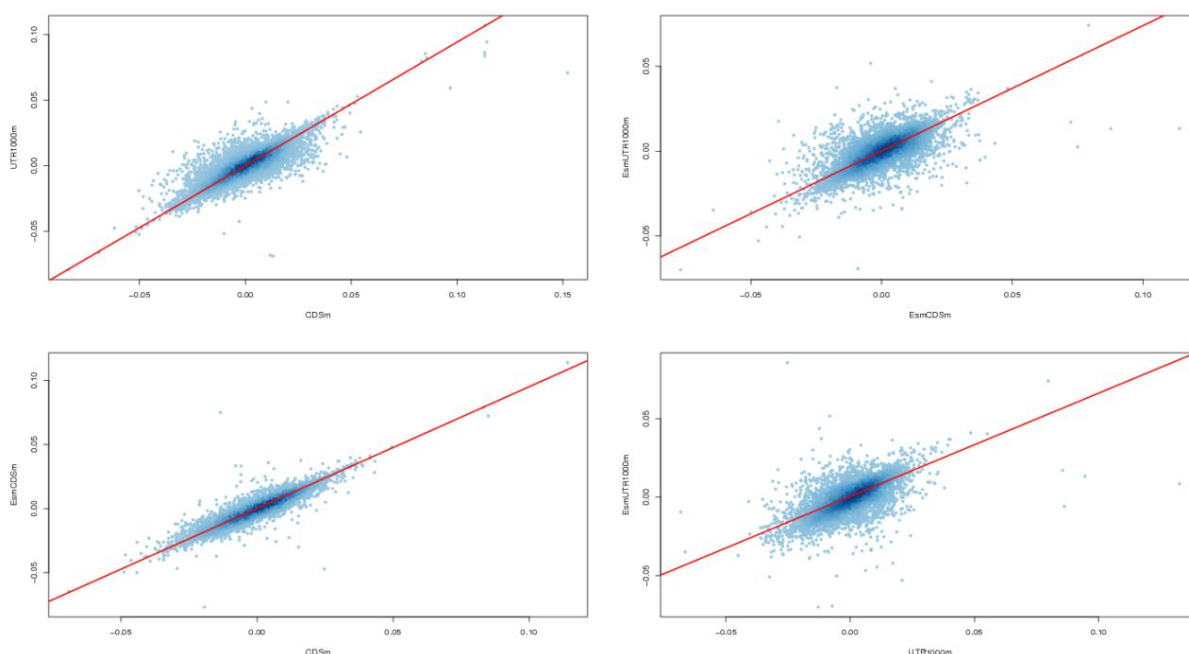


FIGURA 5.10. COMPARAÇÕES DOS VALORES DE DECAIMENTO (COEFICIENTES ANGULARES OU SLOPES) ENTRE OS 4 TIPOS DE MAPEAMENTO.

Mapeamento das sequências apenas contra as regiões codificadoras (mapeamento “CDS”), contra CDS com 5’ UTR (até 70 nt) e 3’ UTR (até 1000 nt) (“UTR”), e o mesmo para mapeamentos únicos apenas no haplótipo *Esmeraldo-like*: “CDS_e” e “UTR_e”. A linha vermelha representa a reta de regressão linear.

FONTE: O autor (2011).

Em seguida também comparamos a média de sequências mapeadas para cada gene entre os quatro tipos de mapeamentos no genoma (FIGURA 5.11). Também foi evidenciada alta correlação entre os resultados dos diferentes métodos, sendo que as distorções são em sua maioria explicáveis pelo maior número de sequências mapeadas, diretamente proporcional ao tamanho em bases da região mapeável (CDS_e < UTR_e < CDS < UTR).

Por esse motivo, ao comparar o número de mapeamentos para o genoma diplóide considerando apenas a CDS *versus* o mapeamento do tipo “UTR”, também diplóide, notamos que para pouquíssimos casos o número de sequências mapeadas

para os genes eram maiores no mapeamento somente na porção codificadora dos genes (CDS), como mostra a FIGURA 5.12 em seguida.

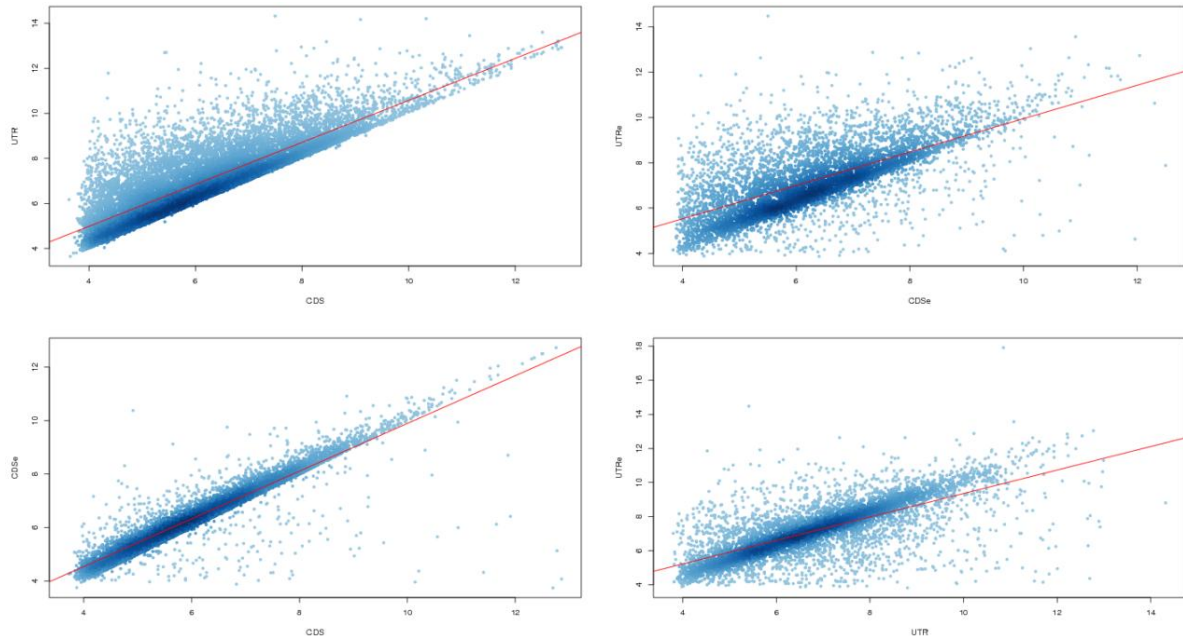


FIGURA 5.11. COMPARAÇÕES DA MÉDIA DE *READS* MAPEADOS ENTRE OS 4 TIPOS DE TAMANHOS DOS GENES.

Mapeamento de *reads* apenas contra as regiões codificadoras (mapeamento “CDS”), contra CDS com 5’ UTR (até 70 nt) e 3’ UTR (até 1000 nt) (“UTR”), e o mesmo para mapeamentos únicos apenas no haplótipo *Esmeraldo-like*: “CDS_{se}” e “UTR_{se}”. A linha vermelha representa a reta de regressão linear. FONTE: O autor (2011).

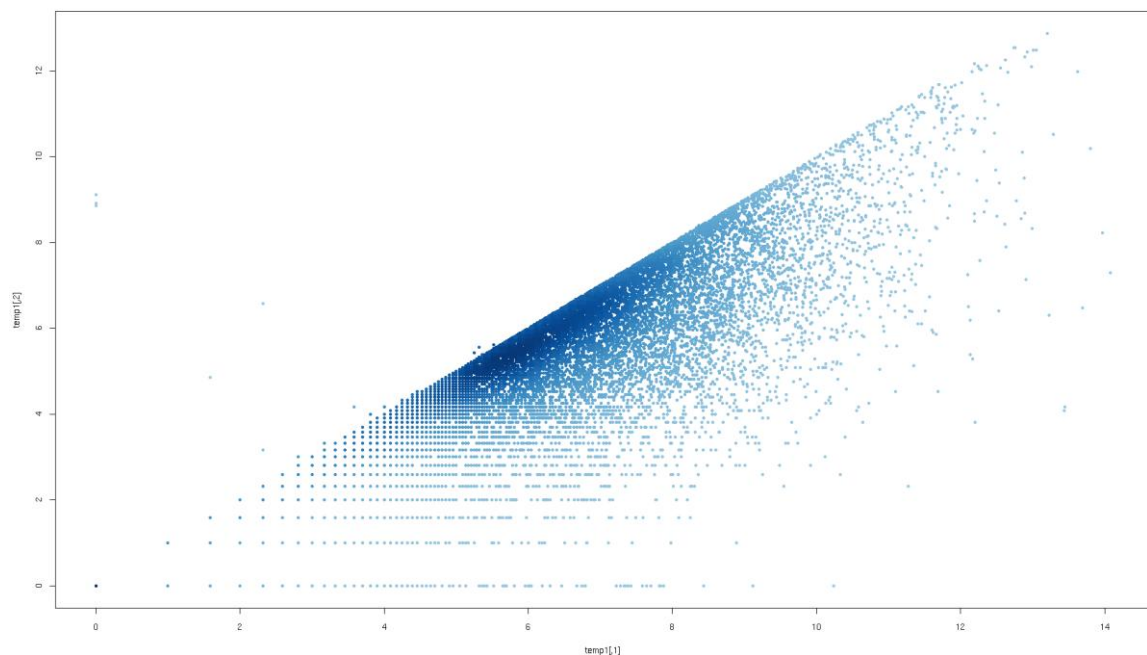


FIGURA 5.12. COMPARAÇÃO DO NÚMERO DE CONTAGENS PRÉ-NORMALIZAÇÃO ENTRE MAPEAMENTOS.

No eixo “x”, o logaritmo na base 2 das contagens para cada gene no mapeamento “UTR”, e no “y” o mesmo para o mapeamento “CDS”. FONTE: O autor (2011).

Portanto, de maneira geral, o padrão de decaimento, incluindo o grau de erro em sua predição, não é fortemente dependente do tipo de mapeamento e sim de propriedades intrínsecas à composição das diferentes amostras analisadas. Sendo assim, escolheu-se o tipo de mapeamento com mais sequências mapeadas (“UTR”), que considera como “gene” até 70 nt de 5’ UTR, a região CDS e até 1000 nt de 3’ UTR, terminando quando um gene a jusante for encontrado antes do limite máximo de 1000 nt. Dessa forma, todas as análises posteriores onde o tipo de mapeamento não for citado foram realizadas utilizando-se esse o tipo de mapeamento no genoma diplóide chamado “UTR”.

5.9 INFLUÊNCIA DA SINEFUNGINA

Com os dados dos pontos 0 e 5 minutos, foi realizada uma análise de expressão diferencial para verificar se a droga sinefungina, colocada no ponto 0 induzia alguma alteração da expressão dos genes, antes da colocação da actinomicina, a partir do ponto 5 minutos. De acordo com a TABELA 5.3, nota-se que, para um limite máximo de FDR de 5% (aceitando-se no máximo 5% de falso positivos no grupo de genes diferencialmente expressos), de 0,86 a 1,10% dos genes mapeados (dependendo do tipo de mapeamento utilizado) se encontram diferencialmente expressos entre a amostra do ponto 0 da série temporal e a amostra de 5 minutos. É interessante notar que o maior número de genes diferencialmente expressos foi identificado na comparação UTR, o qual é provavelmente explicado pela maior amostragem obtida nesse ponto.

Tipo de Mapeamento	FDR < 10%	%	FDR < 5%	%	FDR < 1%	%	FDR < 0.1%	%
CDS - Diplóide	271	1.28	<u>181</u>	0.86	93	0.44	45	0.21
CDS e UTRs - Diplóide	409	1.94	233	1.1	122	0.58	65	0.31
CDS - <i>Esmeraldo-like</i>	132	1.28	87	0.84	41	0.4	20	0.19
CDS e UTRs - <i>Esmeraldo-like</i>	145	1.41	90	0.87	56	0.54	22	0.21

TABELA 5.3. NÚMERO DE GENES DIFERENCIALMENTE EXPRESSOS PARA VÁRIOS LIMITES DE FDR.

FONTE: O autor (2011).

Uma comparação entre os tipos de mapeamentos, com relação ao número de genes com confiança estatística para um limite superior de FDR de 5% em um mapeamento e de 10% no outro mapeamento, se encontra no diagrama de Venn da FIGURA 5.13. Dos 271 genes diferencialmente expressos no mapeamento CDS, 131 também se encontram diferencialmente expressos para um FDR de 10% no mapeamento UTR. Para uma comparação mais restritiva, observe a FIGURA 5.14, a qual mostra quantos dos genes (que passam da linha de corte de FDR menor que 5% em um tipo de mapeamento) também possuem confiança para o mesmo limite de FDR em outro tipo de mapeamento.

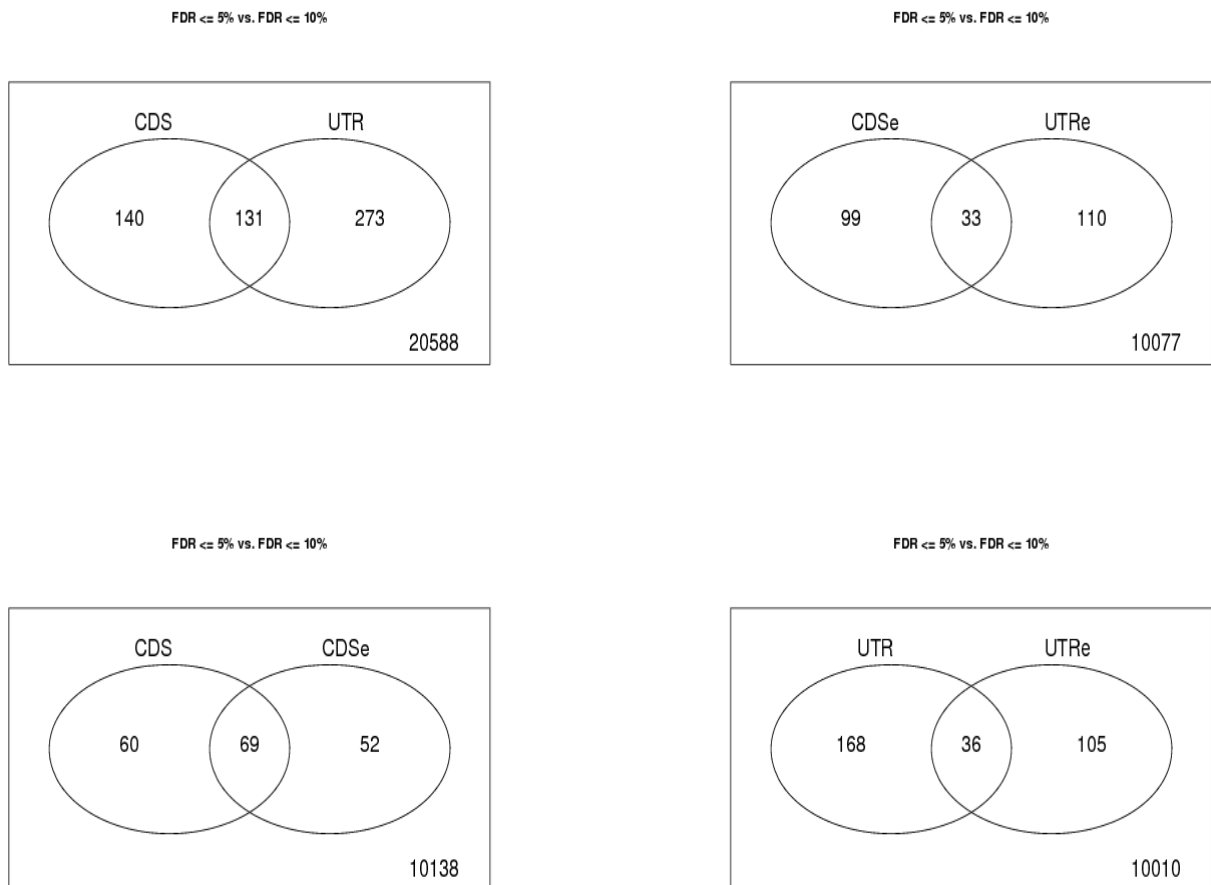


FIGURA 5.13. DIAGRAMA DE VENN DA CONFIANÇA DA EXPRESSÃO DIFERENCIAL EM DIFERENTES MAPEAMENTOS.

No mapeamento da esquerda está o número de genes com $FDR \leq 5\%$ e no grupo da direita os genes com $FDR \leq 10\%$ para seu respectivo tipo de mapeamento dos *reads* no genoma.

FONTE: O autor (2011).

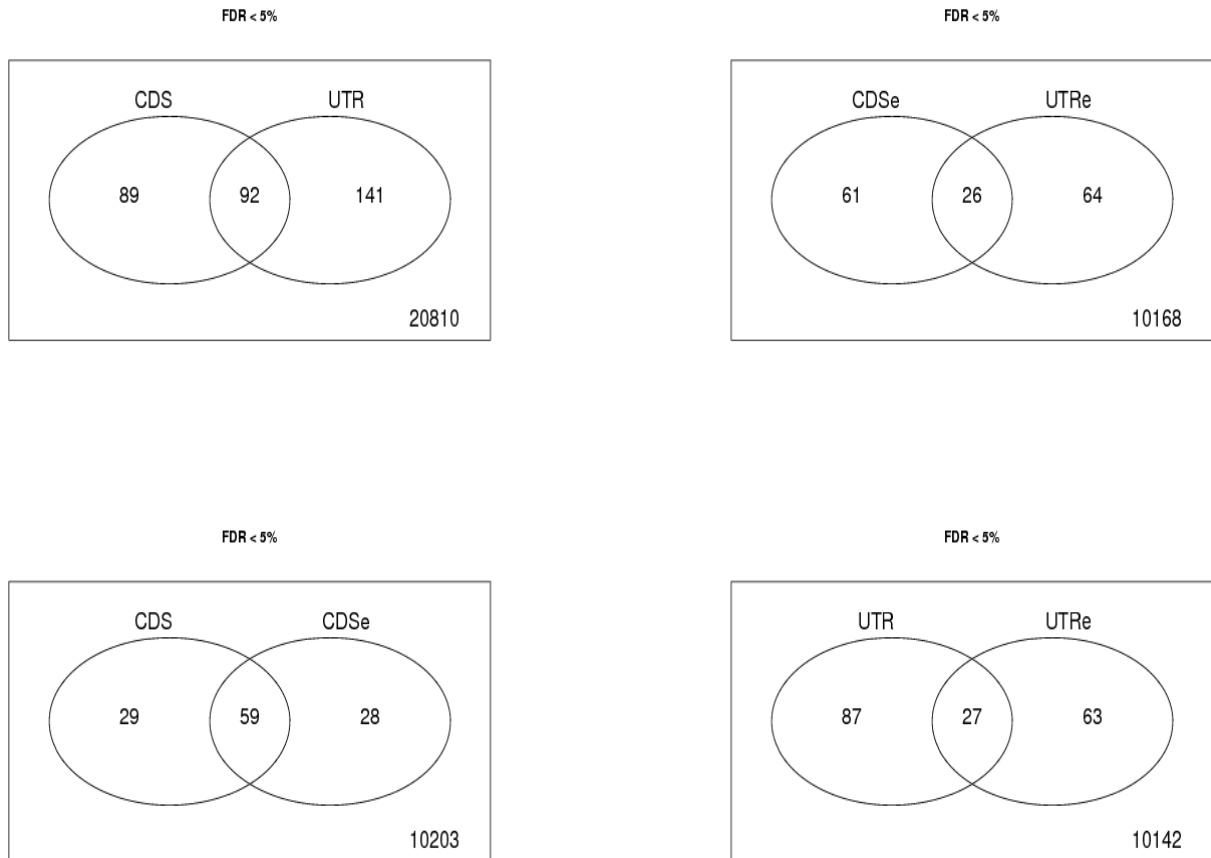


FIGURA 5.14. DIAGRAMA DE VENN DOS GENES DIFERENCIALMENTE EXPRESSOS NOS 4 MAPEAMENTOS.

Em ambos os grupos estão inseridos os genes com FDR menor que 5% para seu respectivo tipo de mapeamento de *reads* no genoma do protozoário.

FONTE: O autor (2011).

A dispersão das contagens dos genes entre os pontos 0' e 5' com relação a sua expressão diferencial no método de mapeamento UTR pode ser visualizada na FIGURA 5.15, que se trata de um gráfico "M versus A" modificado (conhecido por "Smear Plot" ou ainda "Fold Change Plot") que agrupa os genes com contagem zero em uma das réplicas em um grupo separado, a esquerda do gráfico. Os 233 genes diferencialmente expressos são destacados em vermelho e as linhas horizontais em azul representam um *fold-change* de 2 vezes. No eixo "x", o \log_2 da concentração (valor de "A") e no "y" o \log_2 do *fold-change* (valor de "M").

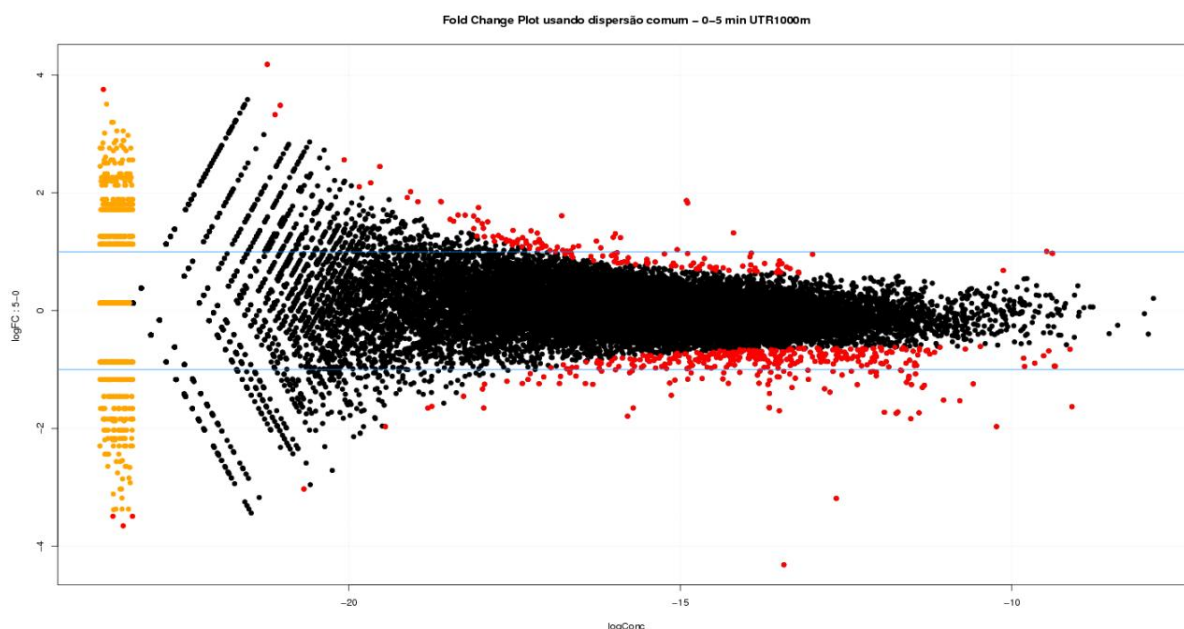


FIGURA 5.15. “SMEAR PLOT” DA EXPRESSÃO DIFERENCIAL ENTRE OS PONTOS 0 E 5 MINUTOS.

Os pontos em vermelho são os genes considerados diferencialmente expressos, os pontos em laranja são os genes que não foram mapeados em uma das réplicas e as linhas azuis delimitam as áreas com aumento (acima) e diminuição (linha abaixo) da expressão de duas vezes ($fold-change = 2$).

FONTE: O autor (2011).

Como exemplo, os 20 primeiros do *ranking* por FDR dos 233 genes diferencialmente expressos com FDR menor do que 5%, suas anotações funcionais e o logaritmo na base 2 do *fold-change* (quantidade de mudança, aumento ou diminuição em número de vezes) estão listados na TABELA 5.4. Nota-se na tabela que a maior parte dos genes listados carece de uma real anotação funcional (proteínas hipotéticas), como é o caso de mais da metade dos genes identificados (anotados) no genoma referência do *Trypanosoma cruzi*, da cepa CL Brener. Esse tipo de falta de atribuição de função dificulta a análise dos resultados dos projetos de genômica funcional que são realizados com esse protozoário flagelado. Também é possível observar que apenas 2 dos 20 genes aumentaram sua expressão 5' após a aplicação da sinefungina no ponto 0'. Ambos com a mesma anotação funcional, a possível “subunidade C11 da RNA Pol III”.

Locus ID	Anotação do gene	Fold Change (Log na base 2)	FDR
TcChr8-S 45	ADP-ribosylation factor, putative	-4,303738	2,34E-64
TcChr8-S 47	ADP-ribosylation factor, putative	-3,182239	1,52E-40
TcChr37-P 524	hypothetical protein	-1,968177	4,80E-17
TcChr35-P 311	hypothetical protein, conserved	-1,835993	1,97E-14
TcChr35-S 323	hypothetical protein, conserved	-1,732957	5,85E-13
TcChr39-P 23	hypothetical protein, conserved	-1,741755	5,85E-13
TcChr39-S 7	hypothetical protein, conserved	-1,719289	9,65E-13
TcChr32-S 157	hypothetical protein, conserved	-1,723042	9,65E-13
TcChr40-S 128	RNA-binding protein, putative	-1,630762	7,44E-12
TcChr40-P 134	RNA polymerase III C11 subunit, putative	1,906160	2,25E-11
TcChr5-P 28	ubiquitin-conjugating enzyme variant Kua homologue (pseudogene), putative	-1,692663	3,42E-11
TcChr40-S 113	RNA polymerase III C11 subunit, putative	1,856439	5,88E-11
TcChr37-P 232	hypothetical protein, conserved	-1,650416	2,06E-10
TcChr37-S 224	hypothetical protein, conserved	-1,643426	2,34E-10
TcChr11-P 78	hypothetical protein, conserved	-1,529666	2,57E-10
TcChr6-P 118	hypothetical protein, conserved	-1,516701	4,10E-10
TcChr21-S 74	hypothetical protein, conserved	-1,772416	3,59E-08
TcChr35-S 253	hypothetical protein, conserved	-1,385886	9,89E-08
TcChr23-P 119	hypothetical protein, conserved	-1,397855	2,08E-07
TcChr39-S 602	acyl-CoA dehydrogenase, putative	-1,302604	3,57E-07

TABELA 5.4. OS 20 GENES MAIS CONFIÁVEIS COM EXPRESSÃO DIFERENCIAL ENTRE OS PONTOS 0' E 5'.

FONTE: O autor (2011).

Para ilustrar todos os 233 genes diferencialmente expressos entre os pontos 0 e 5 minutos, os mesmos foram *clusterizados* por métrica de distância Euclidiana e agrupamento hierárquico UPGMA (FIGURA 5.16). A métrica Euclidiana avalia a distância entre os genes de acordo com a forma do padrão de expressão e a quantidade de contagens. Ou seja, nesse tipo de métrica, a média de sequências mapeadas é considerada e influencia a medida de distância entre o padrão de expressão dos genes, diferentemente do que ocorreria com uma análise cuja métrica de distância fosse baseada em correlação, que é indiferente à magnitude de expressão.

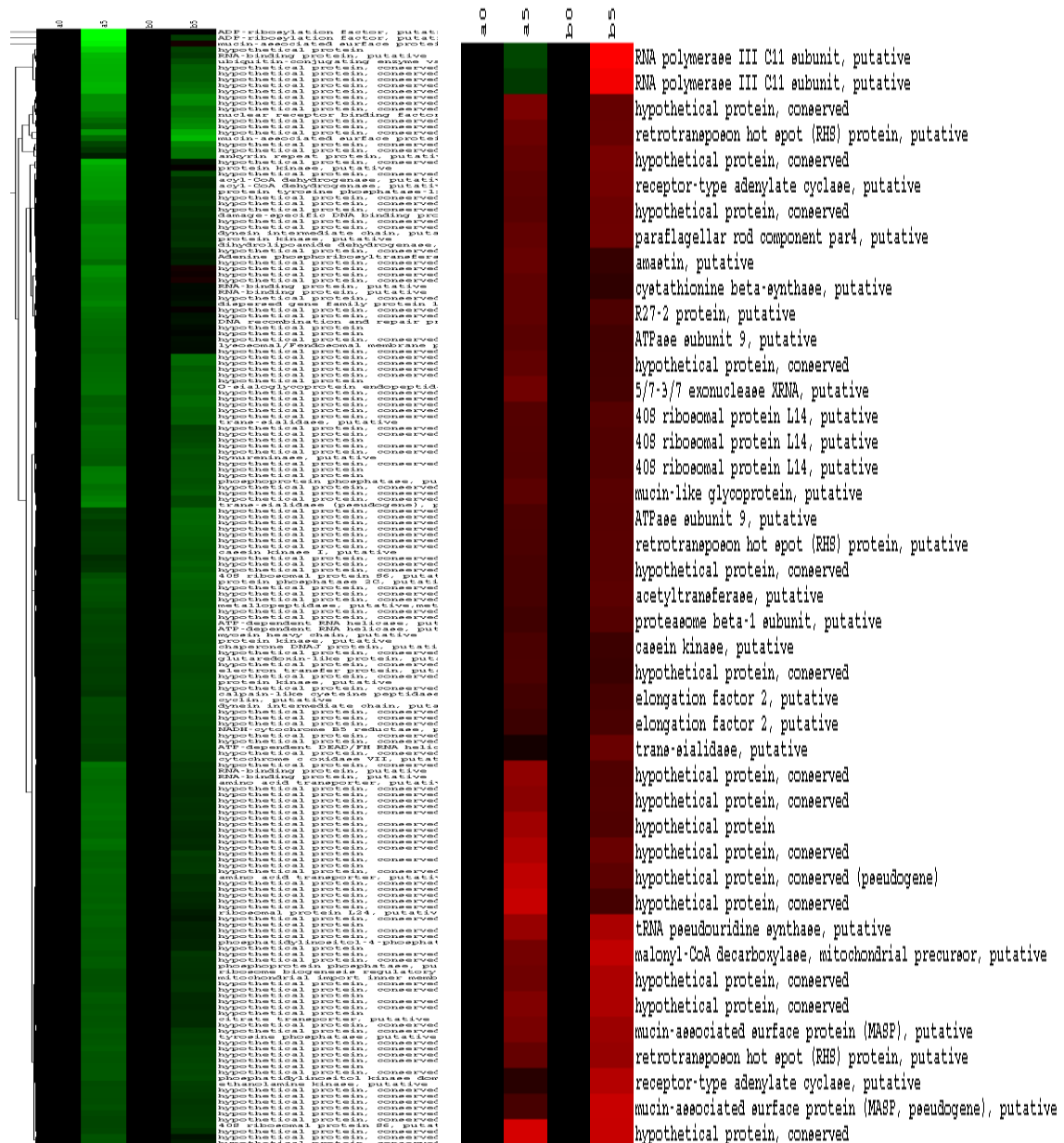


FIGURA 5.16. CLUSTERIZAÇÕES HIERÁRQUICAS DOS GENES DIFERENCIALMENTE EXPRESSOS.

Expressão dos genes (linhas) entre os pontos (colunas 0a, 5a, 0b e 5b, ordenadas a partir da esquerda) após aplicação da droga sinefungina nas duas réplicas. Todos os 233 genes com expressão diferencial com FDR < 5% foram separados em 2 gráficos. Da esquerda para a direita, os 2 gráficos são: 1) *Cluster* dos genes com expressão diminuída com relação aos pontos iniciais (“zero” a e b); 2) *Cluster* dos genes com expressão aumentada em relação aos pontos 0a e 0b. As cores vermelhas simbolizam aumento da expressão gênica aos 5’ em relação ao ponto 0’. E o verde, diminuição. A intensidade das cores informa a intensidade da mudança de expressão gênica.

FONTE: O autor (2011).

5.10 ANÁLISE DO DECAIMENTO

Existe uma grande diferença de complexidade entre uma análise transcriptômica para descobrir quais são os genes diferencialmente expressos entre duas ou mais amostras e uma análise do padrão de expressão geral (no caso, do padrão de decaimento) dos genes. O primeiro modo de análise verifica apenas qual é o grupo de genes cuja expressão foi especialmente alterada mediante duas ou mais situações. Apenas o grupo de genes que ultrapassa um dado limite de corte (que considera o que é expressão diferencial da variação comum não diferencial) é estudado. Já o segundo tipo de análise considera estudar todos os genes em todas as situações ao mesmo tempo, e todos os dados são considerados “informativos”.

Este trabalho possui como foco avaliar as informações sobre a estabilidade dos transcritos provenientes do genoma do *T. cruzi*. Desse modo, todos os dados obtidos derivados da expressão gênica trás consigo informações da taxa de degradação de cada gene.

Em seguida, as contagens de sequências mapeadas para cada gene na série temporal foram regredidas a um modelo linear. Visto que um modelo não-linear (exponencial) seria o modelo esperado (e mais adequado) a ser usado para um experimento de decaimento de moléculas, dois tipos de regressão não-linear foram testados (o método de Lowess com diversos parâmetros e o modelo de regressão exponencial).

Porém, ao compararmos os valores de regressão “ R^2 ” entre os modelos não-lineares aplicados e o modelo linear, não houve diferença significativa na capacidade do modelo matemático explicar os dados. Com o uso de modelos não-lineares a complexidade do tratamento dos dados e de suas análises aumentaria sem necessidade.

Por esse motivo utilizamos o modelo linear, já citado no item anterior, mais simples e direto, facilitando o processo de análise. Essa baixa diferença entre os valores de regressão do modelo linear para os outros modelos pode ser explicada pelo fato de que os tempos de decaimento são precoces (de até 25 minutos). Nessa faixa de tempo, a curva de decaimento é pouco acentuada e bastante íngreme, que a torna muito similar a uma reta. Além disso, há a questão da ocorrência de

relativização entre os dados de decaimento gene a gene em relação a média de decaimento da amostra como um todo, cujo resultado final é a distorção da curva exponencial de decaimento para uma reta de decaimento.

Foi feito o cálculo das regressões lineares para cada gene, considerando as contagens a partir do ponto zero para actinomicina D (o ponto “5 minutos”) até o ponto de 25 minutos. Nisso, observamos que o ponto de 25 minutos destoava muito dos outros pontos da série temporal com relação ao padrão de decaimento e a reprodutibilidade entre as réplicas era menor, de forma geral, para diversos genes. Ou esta seria uma variação inerente a um fenômeno biológico, por exemplo, que em cerca de 25 minutos modificasse a taxa de decaimento da maioria dos genes, ou porque as réplicas de 25 minutos apresentavam problemas técnicos, ou até, porque dentre todos os pontos avaliados, o de 25 minutos é o que tem a maior distorção causada, em relação ao ponto zero, pela inibição da transcrição.

De acordo com o exposto acima, e pelo fato de que o processo de transcrição só começa a ser inibido a partir do ponto 5’, resolvemos excluir os pontos 0’ e 25’ das análises subsequentes.

Para visualizar o decaimento dos genes, fizemos um gráfico com as contagens de *reads* para cada gene em relação à série temporal. Os mesmos foram ordenados decrescentemente com o grau de concordância entre as réplicas, de acordo com o valor da regressão linear (“ R^2 ”). Os dezesseis genes mais confiáveis com transcritos mais instáveis (de rápido decaimento) e os dezesseis genes mais confiáveis com transcritos de menor decaimento (mais estáveis) estão representados nas FIGURA 5.17 e FIGURA 5.18, respectivamente.

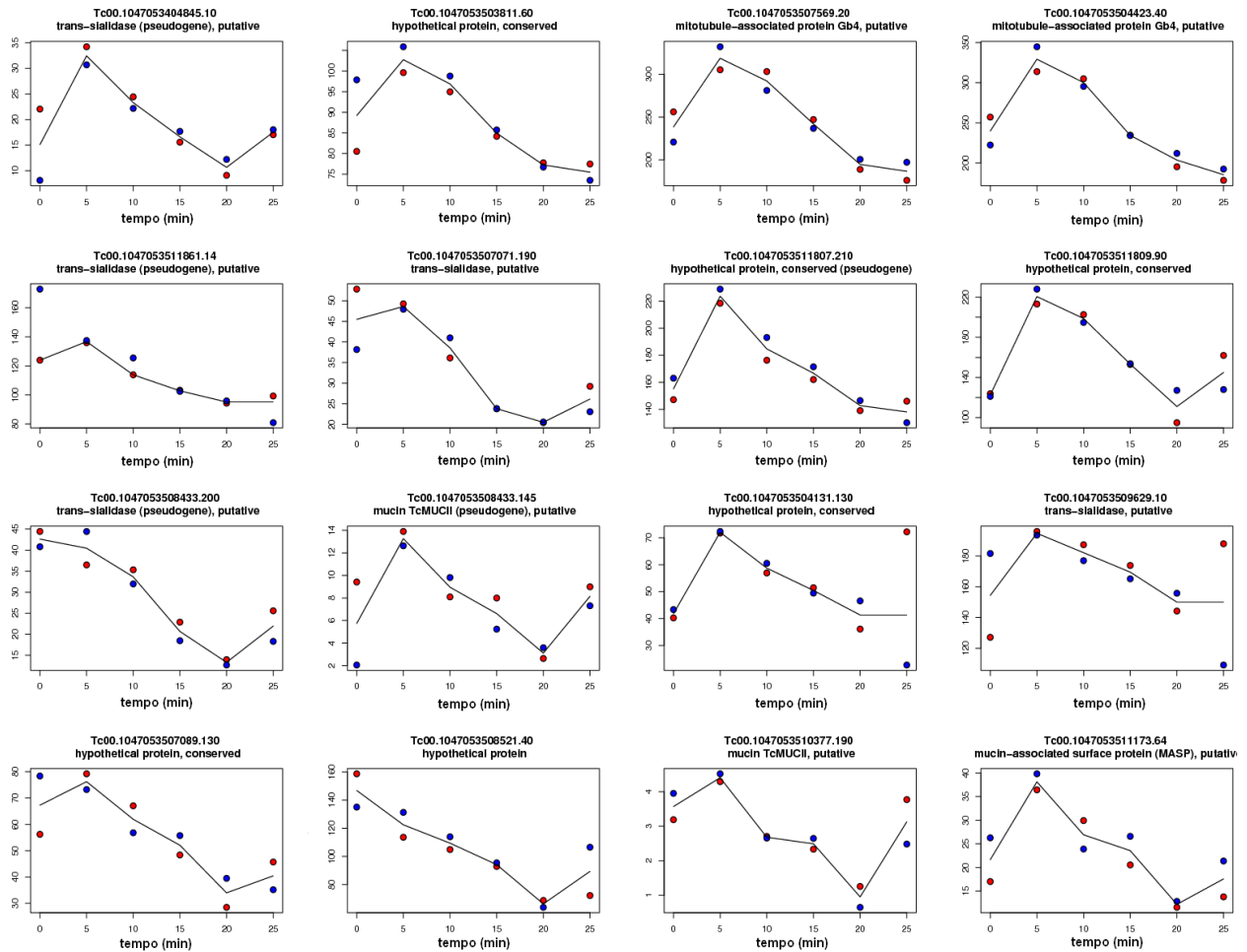


FIGURA 5.17. DECAIMENTO DOS TRANSCRITOS MAIS INSTÁVEIS NA SÉRIE TEMPORAL.

No eixo “y”, está a contagem de reads. Cada cor dos pontos representa uma réplica biológica. Os gráficos dos genes estão ordenados do canto superior esquerdo ao inferior direito, de acordo com seu valor de R^2 na série temporal entre 5 e 20 minutos.

FONTE: O autor (2011).

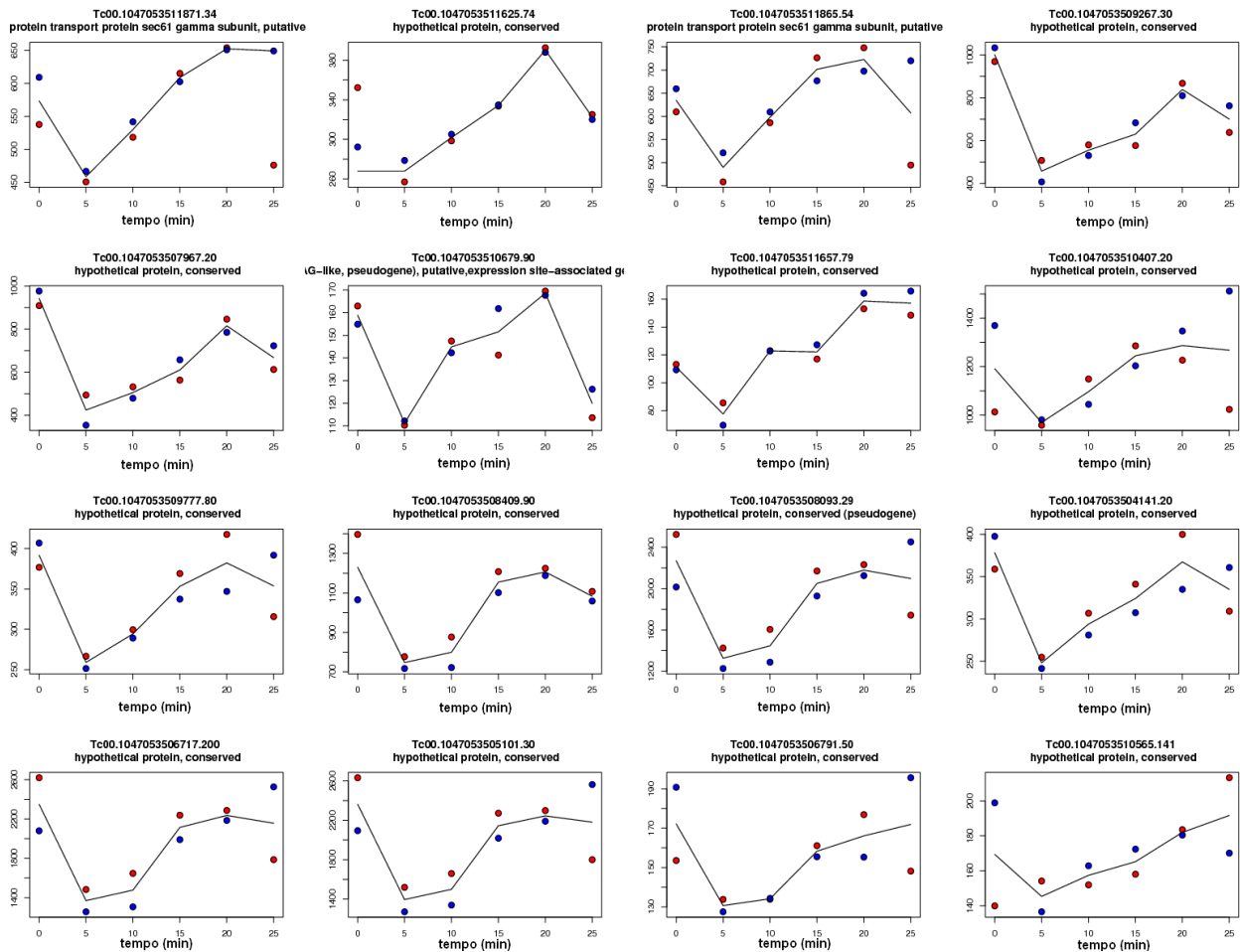


FIGURA 5.18. DECAIMENTO DOS TRANSCRITOS MAIS ESTÁVEIS NA SÉRIE TEMPORAL.

No eixo “y”, está a contagem de *reads*. Cada cor dos pontos representa uma réplica biológica. Os gráficos dos genes estão ordenados do canto superior esquerdo ao inferior direito, de acordo com seu valor de R^2 na série temporal entre 5 e 20 minutos.

FONTE: O autor (2011).

Ao observar qual a distribuição das taxas de decaimento (valor de “*b*”, ou inclinações das retas de regressão linear) entre os genes, fica ressaltado que a maioria dos genes decai de acordo com a média geral de decaimento do transcriptoma (valor de “*b*” igual a zero), como mostra o histograma dos decaimentos na FIGURA 5.19. De certa forma isso é esperado, pois normalmente a regulação da expressão gênica em uma célula deriva de modulações finas pontuais e não de enormes alterações no conteúdo de seu transcriptoma. De tal maneira que o controle da estabilidade *versus* a degradação de determinados tipos de moléculas de mRNA gera uma distribuição em forma de sino, onde a maioria dos genes possui uma taxa de degradação próximo da média do transcriptoma, em concordância com o demonstrado no histograma abaixo.

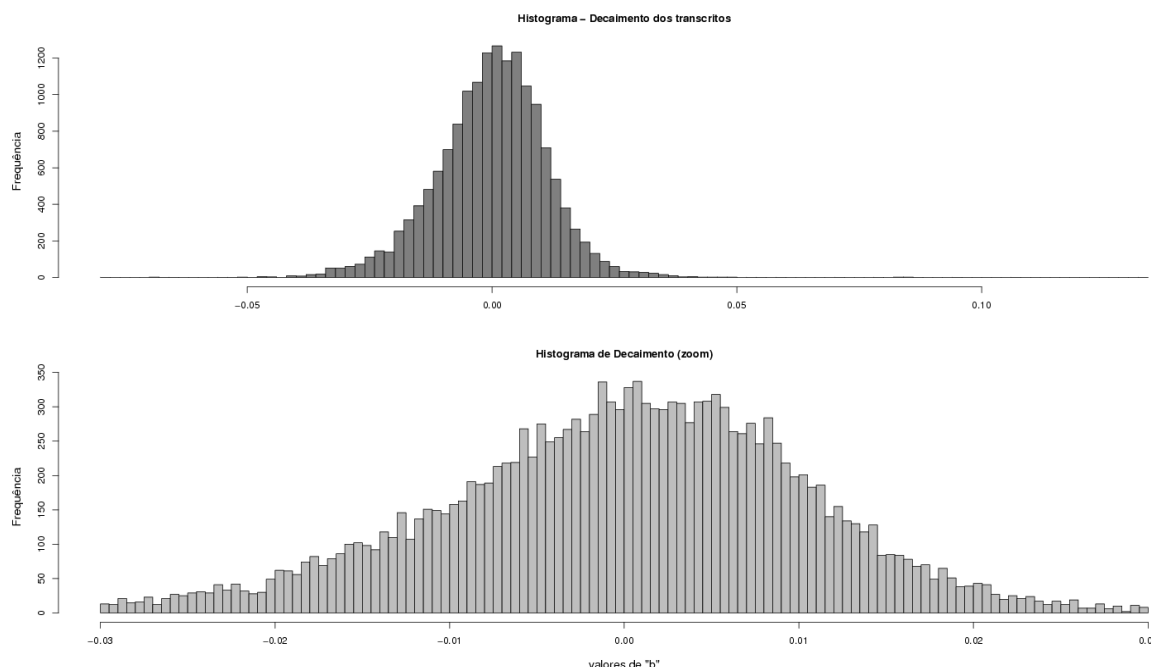


FIGURA 5.19. HISTOGRAMA DAS TAXAS DE DECAIMENTO PARA MAPEAMENTO “UTR”.

Acima: todos os genes, para valores de “b” entre -0,1 e 0,1. Sendo que o valor de “b” igual a zero significa decaimento conforme a média de decaimento da amostra como um todo. Abaixo: apenas os genes com valor de “b” entre -0,03 e 0,03, que representa a porção de maior representatividade do histograma completo, acima.

FONTE: O autor (2011).

5.11 DESCRIÇÕES FUNCIONAIS DOS GENES MAIS ESTÁVEIS E INSTÁVEIS

Em seguida, fizemos uma ordenação dos valores de inclinação das retas de cada gene para saber quais eram os 100 genes mais estáveis (TABELA 5.5) e os 100 que possuíam taxa de degradação mais acentuada (TABELA 5.6). Juntamos as listas das 4 formas de mapeamento no genoma (CDS, UTR, CDSe e UTRe), para que os genes do topo da listagem para cada tipo de mapeamento fossem agrupados. Assim, existem genes com mesmas posições no *rank*, pois se tratam de genes selecionados em tipos de mapeamentos diferentes que foram unidos para confecção das tabelas abaixo.

Essas listas abaixo foram feitas por um *script* em Perl que exclui as famílias multigênicas (RHSs, DGF-1, trans-sialidases, MASPs, mucinas e outras glicoproteínas de superfície como as GP63) e também os genes para proteínas hipotéticas. Já que o primeiro grupo de genes (as famílias multigênicas) é muito

complexo para ser avaliado devido à forma como o genoma referência foi sequenciado e montado e também da maneira pela qual é feito o mapeamento das sequências de *RNA-Seq* no genoma. Por sua vez, o segundo grupo de genes é muito pouco informativo, sem que sejam realizados esforços de anotação muito mais aprimorados, que fogem ao escopo desse trabalho, em possíveis projetos genoma de cepas de *T. cruzi* diferentes da sequenciada atualmente. É possível observar que genes com a mesma notação funcional se encontram nas duas listas, porém, o código identificador (ID) das mesmas é diferente. Apesar de parecer biologicamente incongruente, isso pode estar relacionado com possíveis diferenças de função e de modulação da expressão dos diferentes membros de uma mesma família gênica, o que deverá ser avaliado futuramente.

Rank	ID	Annotation
2	Tc00.1047053510073.10	cysteine peptidase (N-terminal), putative, cysteine peptidase, clan CA, family C1, cathepsin L-like, putative
3	Tc00.1047053507091.50	aspartate carbamoyltransferase, putative
6	Tc00.1047053508903.100	ATPase, putative
7	Tc00.1047053508143.65	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
10	Tc00.1047053508375.5	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
12	Tc00.1047053507937.40	elongation factor 1-gamma (EF-1-gamma), putative
13	Tc00.1047053506941.280	heat shock 70 kDa protein, mitochondrial precursor, putative
16	Tc00.1047053506375.60	tryptophanyl-tRNA synthetase, putative
17	Tc00.1047053508825.10	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
23	Tc00.1047053508767.10	surface protein TolT
24	Tc00.1047053511139.20	40S ribosomal protein S11, putative
28	Tc00.1047053507641.300	chaperonin Hsp60, mitochondrial precursor, groEL protein, heat shock protein 60
29	Tc00.1047053504929.15	casein kinase, delta isoform, putative
34	Tc00.1047053506763.361	syntaxin binding protein (pseudogene), putative
34	Tc00.1047053510841.10	mitotic centromere-associated kinesin (MCAK, pseudogene), putative
38	Tc00.1047053509717.152	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
40	Tc00.1047053506001.111	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
41	Tc00.1047053511473.9	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
42	Tc00.1047053510627.179	syntaxin binding protein 1 (pseudogene), putative
47	Tc00.1047053507559.80	deoxyribose-phosphate aldolase, putative
48	Tc00.1047053511553.71	syntaxin binding protein (pseudogene), putative
49	Tc00.1047053506763.361	syntaxin binding protein (pseudogene), putative
54	Tc00.1047053508321.11	histone H2A, putative
54	Tc00.1047053508143.65	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
56	Tc00.1047053508401.90	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
58	Tc00.1047053506603.29	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
61	Tc00.1047053506271.20	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
64	Tc00.1047053511621.90	esterase, putative
65	Tc00.1047053508321.21	histone H2A, putative
68	Tc00.1047053508375.50	dihydroorotate dehydrogenase, putative
69	Tc00.1047053504149.210	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
75	Tc00.1047053509233.190	ribosomal protein S25, putative
76	Tc00.1047053506715.20	metallopeptidase, putative, metallo-peptidase, clan MP, family M67, putative
78	Tc00.1047053511417.90	phosphate-repressible phosphate permease, putative
81	Tc00.1047053510755.60	ADP-ribosylation factor 1, putative
98	Tc00.1047053509753.65	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative

TABELA 5.5. TRANSCRITOS MAIS ESTÁVEIS, INDEPENDENTE DO TIPO DE MAPEAMENTO.

Da esquerda para a direita, cada coluna representa: (1) posição no *ranking*, (2) código único do gene e (3) anotação funcional automática (derivada do projeto genoma) atribuída ao gene.

FONTE: O autor (2011).

Rank	ID	Annotation
1	Tc00.1047053510339.100	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative
1	Tc00.1047053507641.290	chaperonin Hsp60, mitochondrial precursor, groEL protein, heat shock protein 60
3	Tc00.1047053510693.310	syntaxin binding protein (pseudogene), putative
4	Tc00.1047053510279.40	90 kDa surface protein, putative, serine-alanine- and proline-rich protein, putative
6	Tc00.1047053508269.73	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
14	Tc00.1047053503421.14	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
15	Tc00.1047053509723.20	glycine dehydrogenase (pseudogene), putative
16	Tc00.1047053507145.31	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
19	Tc00.1047053509391.10	calcium-binding protein, putative
20	Tc00.1047053508193.60	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase, putative
20	Tc00.1047053511351.30	protein kinase, putative
21	Tc00.1047053511807.3	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
23	Tc00.1047053506179.5	ferric reductase, putative
24	Tc00.1047053507957.36	syntaxin binding protein (pseudogene), putative
25	Tc00.1047053509705.10	cellulosomal scaffoldin anchoring protein, putative
27	Tc00.1047053504427.260	kinesin, putative
28	Tc00.1047053506967.27	syntaxin binding protein (pseudogene), putative
29	Tc00.1047053507957.36	syntaxin binding protein (pseudogene), putative
30	Tc00.1047053504213.80	threonine synthase, putative
31	Tc00.1047053511621.70	ribosomal protein L11, putative
32	Tc00.1047053508221.730	90 kDa surface protein, putative
32	Tc00.1047053507959.323	syntaxin binding protein (pseudogene), putative
33	Tc00.1047053505365.50	beta galactofuranosyl glycosyltransferase (pseudogene), putative
34	Tc00.1047053510361.274	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
34	Tc00.1047053503611.20	ribosomal protein L24, putative
36	Tc00.1047053510025.7	syntaxin binding protein (pseudogene), putative
37	Tc00.1047053504155.200	syntaxin binding protein (pseudogene), putative
43	Tc00.1047053509959.198	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
44	Tc00.1047053510025.7	syntaxin binding protein (pseudogene), putative
46	Tc00.1047053510345.9	protein kinase, putative
47	Tc00.1047053511019.54	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
47	Tc00.1047053510307.180	glycine dehydrogenase [decarboxylating] (pseudogene), putative, glycine dehydrogenase [decarboxylating], putative, frameshift
48	Tc00.1047053504109.195	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
49	Tc00.1047053510525.80	histone H2A, putative
50	Tc00.1047053511599.85	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
50	Tc00.1047053509959.198	ATP-dependent DEAD/FH RNA helicase (pseudogene), putative
51	Tc00.1047053509391.20	calcium-binding protein, putative
54	Tc00.1047053504155.110	syntaxin binding protein (pseudogene), putative
55	Tc00.1047053508221.451	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
56	Tc00.1047053509581.20	beta galactofuranosyl glycosyltransferase, putative
60	Tc00.1047053507145.31	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
65	Tc00.1047053511613.100	elongation factor 1-gamma (EF-1-gamma), putative
66	Tc00.1047053506129.60	beta galactofuranosyl glycosyltransferase, putative
68	Tc00.1047053509265.23	N-acetyltransferase complex ARD1 subunit (pseudogene), putative
70	Tc00.1047053504163.50	fructose-bisphosphate aldolase, glycosomal, putative
70	Tc00.1047053506743.110	protein tyrosine phosphatase, putative
71	Tc00.1047053507237.110	elongation factor 1-gamma (EF-1-gamma), putative
71	Tc00.1047053504929.10	casein kinase, delta isoform, putative
72	Tc00.1047053508671.20	serine carboxypeptidase (CBP1), putative, serine peptidase, Clan SC, Family S10, putative
75	Tc00.1047053511255.681	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
80	Tc00.1047053508375.30	aspartate carbamoyltransferase, putative
81	Tc00.1047053506717.219	GMP synthase (pseudogene), putative
83	Tc00.1047053507641.180	endochitinase, putative
85	Tc00.1047053507641.104	small GTPase, putative
86	Tc00.1047053508121.43	elongation factor 1-gamma (EF-1-gamma, pseudogene), putative
89	Tc00.1047053507641.210	T-lymphocyte triggering factor, putative

TABELA 5.6. TRANSCRITOS MENOS ESTÁVEIS, INDEPENDENTE DO TIPO DE MAPEAMENTO.

Da esquerda para a direita, cada coluna representa: (1) posição no *ranking*, (2) código único do gene e (3) anotação funcional automática (derivada do projeto genoma) atribuída ao gene.

FONTE: O autor (2011).

Adicionalmente, cabe ressaltar que a anotação do genoma é feita com base em similaridade de sequência das ORFs (quadros de leitura aberta, do inglês “*Open Reading Frames*”) encontradas com genes de outros organismos em bancos de dados. As ORFs são as regiões encontradas num genoma recém montado que detêm algumas características das regiões codificadoras de proteínas. Estas são comparadas a bancos de dados de sequências conhecidas e a função dos novos genes definidos é inferida com base em características predominantes como domínios funcionais e similaridade de estrutura primária com genes já anotados. Contudo, as pequenas diferenças como SNPs (“*Single Nucleotide Polymorphism*”,

as mutações pontuais) e em muitos casos regiões maiores de dissimilaridade ou até porções deletadas/inseridas são desconsideradas. Outro fator complicador é o próprio método de mapeamento de *reads* no genoma referência, que pode contar como leitura mapeada para mais de um gene (em regiões de igualdade de sequências), no caso dos mapeamentos nos 2 haplótipos.

É possível notar que existe uma grande diferença no número de genes selecionados entre os 100 mais estáveis e os 100 de rápido decaimento, respectivamente, na Tabela 5.5 e na Tabela 5.6. A segunda possui quase o dobro de genes que não fazem parte das famílias multigênicas e não são proteínas hipotéticas. Após verificar o tipo de função atribuída a tais genes, percebe-se uma sobre-representação para funções relacionadas à tradução de proteínas, segundo uma análise de enriquecimento para termos de ontologia gênica (termos “GO”) relacionados aos genes da lista. Dessa forma, seria possível relacionar os transcritos maduros que codificam proteínas envolvidas com o mecanismo de tradução com altas taxas de decaimento. Já o grupo de transcritos mais estáveis, não mostrou favorecer nenhuma categoria funcional específica.

A partir dessas listas (porém, agora completas, sem o filtro que retira os genes das famílias multigênicas citadas e os genes para proteínas hipotéticas), fizemos análises para avaliar as relações funcionais entre os genes com transcritos mais estáveis e os que possuem mRNAs que são degradados mais rapidamente que o decaimento médio da população de transcritos. Para tanto, os genes ordenados, de forma decrescente, pelo seu coeficiente de inclinação da regressão linear (que representa a taxa de decaimento) foram divididos em 20 grupos de 50, 100, 200 e 500 genes. Sendo metade dos grupos de genes com maior decaimento e metade para os de maior estabilidade. Tanto para os genes com maior taxa de decaimento quanto para os mais estáveis, os grupos de 1 a 10 ordenados representam os 10 grupos (com 50, 100, 200 ou 500 integrantes) de genes com maior intensidade (sendo o grupo 1 o mais intenso) da inclinação da reta de regressão, ascendente (grupos “*top*”) ou descendente (grupos “*bottom*”). Por exemplo, para as separações de grupos com 500 genes, o grupo “*top1*” inclui a lista dos genes de maior taxa de decaimento do 1º até o 500º gene, o grupo “*top2*” do 501º ao milésimo gene, e assim sucessivamente. O mesmo para os grupos “*bottom*”, porém no sentido contrário da lista ordenada, da base para o meio. Tais grupos foram submetidos a exames do (i) padrão de expressão gênica no ciclo de vida do

parasito, de acordo com os dados obtidos por nosso grupo usando microarranjo de DNA (PAVONI, 2005), (ii) tamanho do gene e (iii) enriquecimento para termos de ontologia gênica (“termos GO”, do inglês “*Gene Ontology*”) utilizando a ferramenta *Blast2GO* (CONESA et al., 2005) ao invés de avaliar manualmente cada anotação da lista de genes.

De acordo com o primeiro teste realizado, não fica evidente uma forte correlação dos grupos gênicos “*top*” ou “*bottom*” como um todo e sua expressão diferencial entre os estágios (os tipos celulares) do ciclo de vida do *T. cruzi*. Porém, notou-se que, para alguns genes, existe uma grande relação do padrão de decaimento com sua expressão no ciclo de vida. Logo, a regulação da expressão gênica que determina a diferenciação celular do parasito em seu ciclo de vida, para alguns genes, pode ser explicada (ao menos parcialmente) pela estabilidade de seus transcritos maduros, como demonstra a *clusterização* hierárquica do padrão de expressão durante o ciclo de vida para os 500 genes do grupo *top1* (FIGURA 5.20). Nessa figura, genes mais estáveis se encontram mais expressos em epimastigotas. Já na FIGURA 5.21, alguns dos genes com rápido decaimento dos grupos *top* de 100 genes estão com expressão diminuída nessa mesma forma celular.

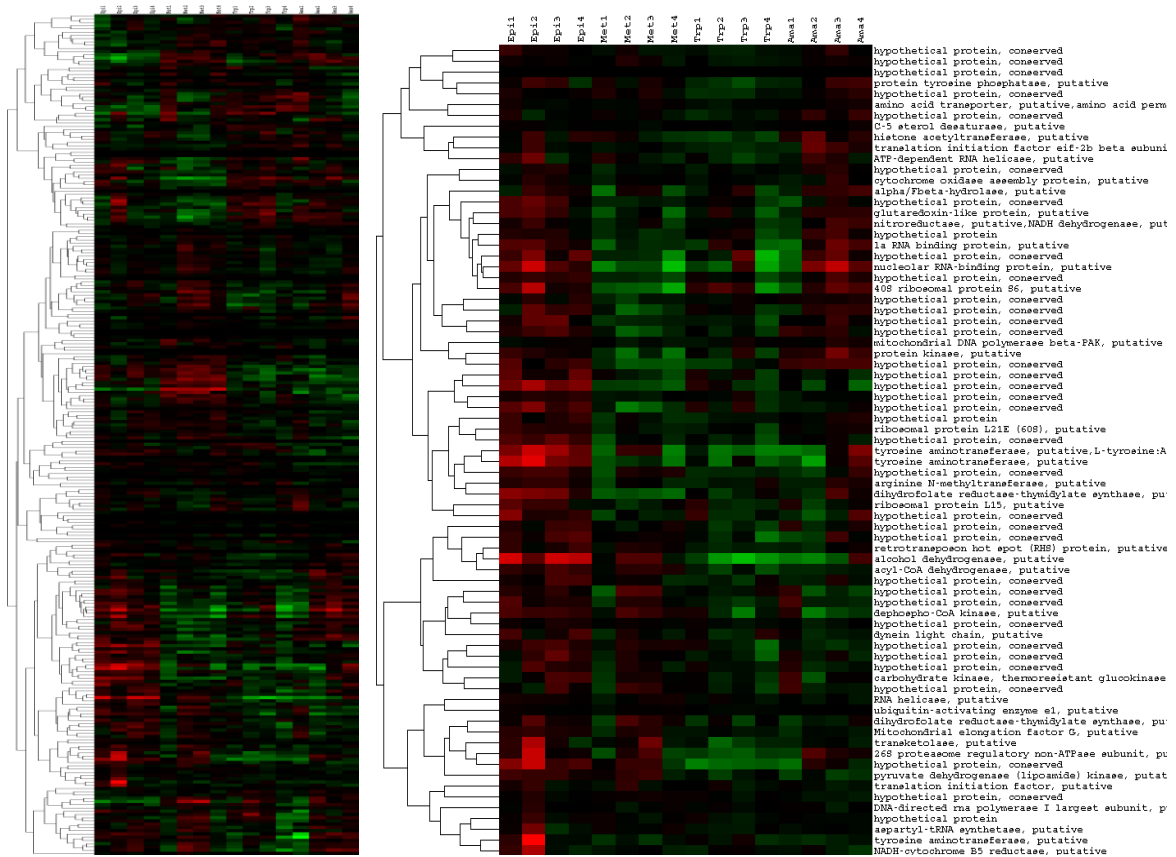


FIGURA 5.20. PADRÃO DE EXPRESSÃO NO CICLO DE VIDA DOS 500 TRANSCRITOS MAIS ESTÁVEIS.

Ao lado esquerdo, as linhas da clusterização são os 500 genes com os transcritos de menor taxa de decaimento e as 16 colunas são referentes à expressão desses genes no ciclo de vida por dados de microarranjos com 4 réplicas para cada fase do ciclo (em ordem, epimastigotas, tripomastigotas metacíclicos, tripomastigotas e amastigotas). No gráfico à direita, uma ampliação de um grupo dos 500 genes mais estáveis cuja expressão se encontra aumentada em epimastigotas em relação às outras fases do ciclo de vida nos dados de microarranjo. Tons de verde simbolizam diminuição na expressão gênica e tons de vermelho aumento da mesma, sendo que a intensidade da expressão está representada na saturação das cores. FONTE: O autor (2011).

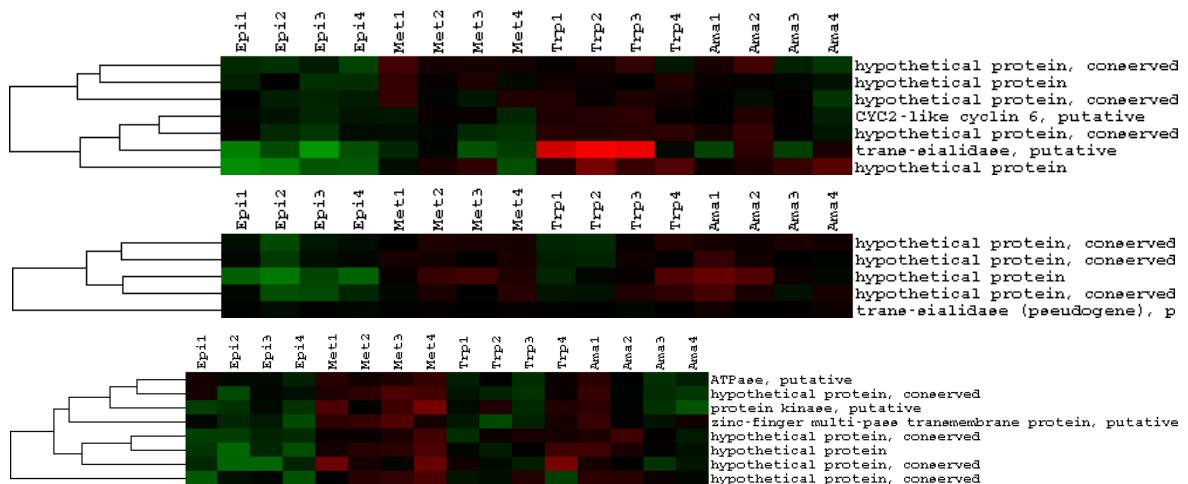


FIGURA 5.21. PADRÃO DE EXPRESSÃO NO CICLO DE VIDA DE ALGUNS GENES DE DECAIMENTO ACENTUADO.

Em verde, diminuição da expressão e em vermelho, aumento da mesma, como na figura anterior. FONTE: O autor (2011).

Seguindo para o próximo teste, para ver se existia alguma relação entre o tamanho dos genes e sua velocidade de decaimento, comparamos essas duas variáveis. Já que um RNA com mais nucleotídeos poderia, ao acaso, demorar mais para ser degradado apenas por ter tamanho maior. Como mostra a FIGURA 5.22, não existe relação evidente entre o tamanho do gene e sua taxa de degradação.

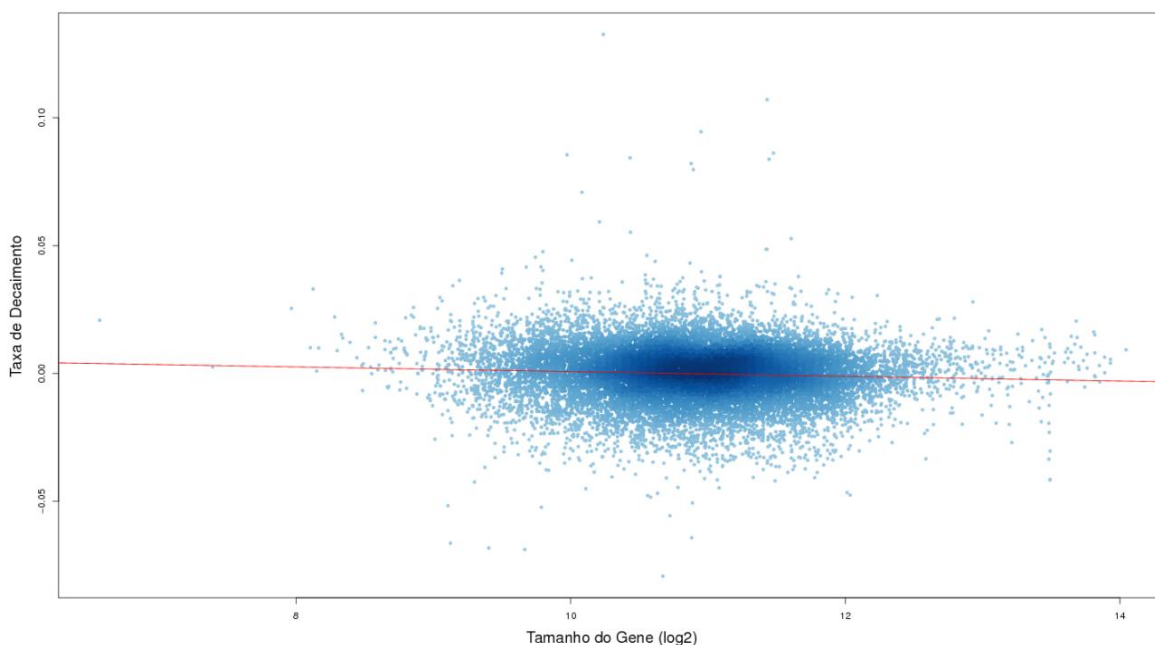


FIGURA 5.22. COMPARAÇÃO ENTRE O TAMANHO DO GENE E A VELOCIDADE DE DEGRADAÇÃO DE SEU TRANSCRITO.

Os pontos e a linha de tendência dos mesmos mostram ausência de relação entre essas duas grandezas. A taxa de decaimento está representada pelo valor de “*b*”.

FONTE: O autor (2011).

O próximo teste, de enriquecimento para termos “GO” se encontra na FIGURA 5.23. Nela, pode ser observado que o grupo dos 500 genes mais estáveis possui alguns termos mais representados (com FDR menor que 5%) que na distribuição dos mesmos no genoma do *T. cruzi*, tais como “organização do citoesqueleto”, “organização de componente celular” e “organização de organelas”, além de outros termos que são termos mais genéricos que incluem esses três. Isso possibilita a existência de uma provável relação entre estabilidade dos transcritos e modificações do citoesqueleto e de genes envolvidos com organelas (principalmente as não envoltas por membranas, como ribossomos, proteassomos, *p-bodies*, entre outros complexos macromoleculares). Os grupos com representação abaixo do encontrado no grupo referência (o genoma) são ignorados por uma questão de

amostragem, já que, do tamanho total (500) apenas 159 genes possuíam um termo “GO” associado e foram enviados para análise de enriquecimento.

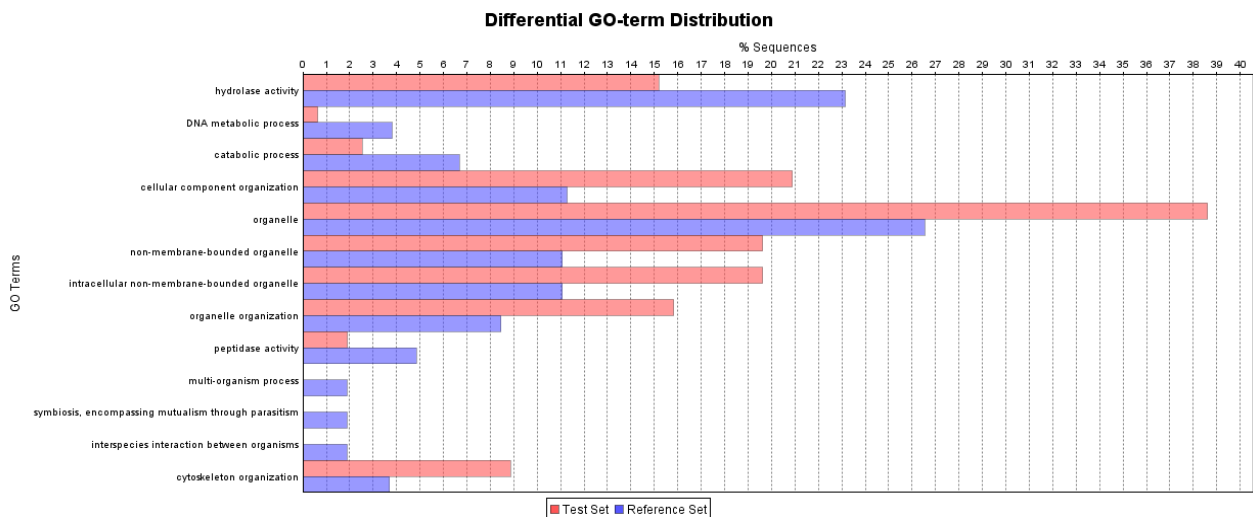


FIGURA 5.23. REPRESENTATIVIDADE DOS TERMOS “GO” NO GRUPO DOS 500 TRANSCRITOS MAIS ESTÁVEIS.

Em vermelho o conjunto teste (os 500 transcritos mais estáveis) e em azul o conjunto referência.

FONTE: O autor (2011).

5.12 ANÁLISE DA REPRODUTIBILIDADE DO PADRÃO DE DECAIMENTO

Para cada um dos genes mapeados foi calculada a correlação de Pearson entre suas contagens de *reads* das duas réplicas de toda a série de tempo de exposição às drogas. Posteriormente, para ver se existia alguma relação direta entre a correlação de Pearson entre as réplicas e alguma característica inerente aos genes, comparamos a correlação calculada com o número total de sequências mapeadas, o tamanho e o decaimento de cada gene. Não foi percebida nenhuma relação forte entre a reprodutibilidade do padrão de decaimento dos genes entre as réplicas e a contagem de mapeamentos, o tamanho considerado para cada gene ou a estabilidade de seus mRNAs, segundo demonstrado na FIGURA 5.24.

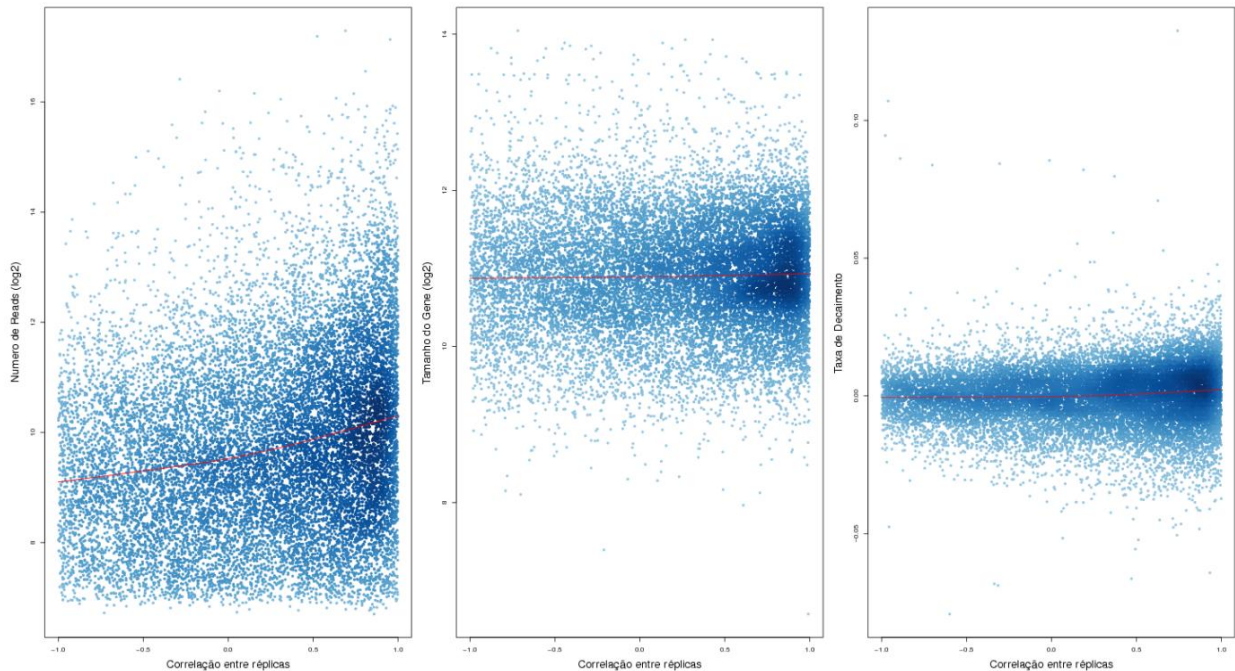


FIGURA 5.24. RELAÇÃO ENTRE A REPRODUTIBILIDADE DO PADRÃO DE EXPRESSÃO DOS GENES ENTRE AS RÉPLICAS DA SÉRIE TEMPORAL E OUTRAS CARACTERÍSTICAS DOS GENES.

Da esquerda para a direita, a comparação da correlação entre as réplicas (eixo “x”) e (1) número de *reads*, (2) tamanho do gene e (3) taxa de decaimento, respectivamente. As linhas de tendência vermelhas são as regressões não lineares (método de *lowess*) dos pontos dos gráficos.

FONTE: O autor (2011).

Existe, todavia, uma leve tendência de aumento da reprodutibilidade com relação ao aumento do número total de *reads* mapeados para cada gene. Percebe-se também, que a maior parte dos genes é bastante reprodutível, pois a densidade maior de pontos se encontra na região próxima a correlação de Pearson de 1 (total correlação positiva). Também não foi encontrada nenhuma relação dos genes com coeficiente de correlação de Pearson entre réplicas maiores que 0,9 ou menores que -0,9 com enriquecimento para termos de ontologia gênica pela ferramenta *Blast2GO*.

6 CONCLUSÃO

O presente trabalho consiste na primeira avaliação em larga escala do padrão de decaimento de mRNAs em um protozoário com controle predominantemente pós-transcricional de sua expressão gênica; também consiste no primeiro trabalho avaliando o decaimento pela técnica de RNA-Seq em um organismo patogênico.

Foi possível avaliar com alto grau de precisão e confiabilidade, pelo uso da técnica de RNA-Seq e pela utilização de um conjunto relativamente amplo de pontos temporais precoces, o padrão de decaimento de mRNAs em *Trypanosoma cruzi* cepa Y.

De acordo com os resultados obtidos, verificou-se que a complexidade da análise dos padrões de decaimento é mais alta do que esperávamos no início desse projeto, havendo diversos fatores que influenciam negativamente na determinação de um padrão de decaimento. Esses fatores são: baixo grau de mapeabilidade (30%), famílias multigênicas presentes em *T. cruzi*, genoma sequenciado com montagem e anotação complexa, influência do grau de decaimento médio na determinação do padrão de decaimento gene-específico, mudança desse padrão de decaimento médio das amostras durante o experimento e alteração do padrão de decaimento de cada mRNAs ao longo do processo (por regulação da expressão).

Por outro lado, a diversidade de padrões observados é menor do que esperado a princípio. O resultado obtido foi a de que a maioria dos mRNAs de *T. cruzi* seguem um padrão de decaimento médio comum e, portanto, seu valor de decaimento (*slope*, parâmetro “*b*” da equação linear) é próximo a zero.

No entanto, pela abordagem em larga escala, esses dados representam a real complexidade biológica envolvida nesse modelo de decaimento (uso das drogas inibitórias), que é ampliada pela complexidade inerente à análise do transcriptoma de *T. cruzi*.

Conseguimos evidenciar alguns grupos que possuem estabilidade de seus transcritos em epimastigotas, os quais parecem estar envolvidos com etapas do processo de tradução do mRNA em proteínas (ribossomos), com complexos

macromoleculares do tipo RNPs (*p-bodies*, proteassomos etc.) e com organização de organelas e citoesqueleto.

Existe a possibilidade, apesar de muito baixa, de as drogas não terem agido da forma esperada visto que existe uma grande variação do grau de decaimento de muitos genes (conforme visto na Figura 5.18, por exemplo), como se para muitos genes não houvesse um claro padrão de decaimento. Esta hipótese será testada em trabalhos futuros com novos experimentos.

7 PERSPECTIVAS

Os dados resultantes estão sendo integrados às plataformas do ICC de análise de genômica funcional de *T. cruzi* (CruziGeneDB e KinetoDB) e analisados de acordo com diversos programas, de uso geral ou desenvolvidos internamente para a realização de diferentes tarefas de *data mining*.

Almejo também testar mais hipóteses que expliquem a baixa mapeabilidade das sequências obtidas no equipamento, já que para outros experimentos de *RNA-Seq* realizados do instituto se obteve, em média, mapeamento para 80% das sequências geradas.

Pretendemos refazer os experimentos com a cepa Dm28c e comparar seu sequenciamento com este realizado com a cepa Y, para avaliar as diferenças ou similaridades entre as cepas e para seguir ampliando a base de dados do projeto Reguloma. A comparação entre essas duas cepas faz parte de um projeto maior de identificação de diferenças ômicas entre Y e Dm28c, que são filogeneticamente distantes e que apresentam perfis de crescimento e de diferenciação, entre outros, muito diferentes.

Ao término do sequenciamento do genoma da cepa Dm28c (iniciado no ano passado em uma colaboração do ICC com o LNCC – Laboratório Nacional de Computação Científica), tenho o intuito de mapear esse conjunto de sequências dos experimentos de decaimento no genoma de Dm28c. Análises preliminares de nosso grupo indicam que o grau de mapeabilidade aumenta de 5% a 10% com o uso de

sequências de Dm28c (de 80% para 85-90%), bem como um aumento generalizado no escore médio de mapeamento.

Além disso, esse trabalho já possui um projeto de continuidade para meu processo de doutoramento que propõe a mesma avaliação em larga escala de degradação de mRNAs por meio de ensaios de inibição da transcrição avaliando diversas situações do ciclo de vida do protozoário parasita em questão e não apenas nas formas epimastigotas. A diferenciação das formas epimastigotas em tripomastigotas metacíclicos pode ser realizada *in vitro* sob condições quimicamente definidas (CONTRERAS *et al.*, 1985; CONTRERAS *et al.*, 1988) e fará parte dos próximos estudos de decaimento de RNAs. As situações do ciclo de vida serão: diferentes formas celulares (amastigota, tripomastigota, epimastigota, epimastigota com 24 horas de diferenciação durante a metaciclogênese, tripomastigota metacíclico e um ponto intermediário no processo de epimastigogênese), cinco condições de estresse ambiental (variações de temperatura, de pH; nutricionais e estresse oxidativo) e depleção de alguns genes regulatórios individualmente por técnica de nocaute gênico.

O objetivo do meu projeto de doutoramento é ampliar fortemente a base de decaimento de mRNA presente em nossos bancos de dados, bem como entender como dinâmicas de modulação no nível de expressão de proteínas reguladoras do nível de mRNAs influenciam nos resultados finais obtidos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABUIN, G.; FREITAS-JUNIOR, L.H.; COLLI, W.; ALVES, M.J.; SCHENKMAN, S. Expression of trans-sialidase and 85-kDa glycoprotein genes in *Trypanosoma cruzi* is differentially regulated at the post-transcriptional level by labile protein factors. **The Journal of Biological Chemistry**, v. 274, n. 19, p. 13041-13047, 1999.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. (eds). trad. VEIGA, A. B. G. *et al.* **Biologia Molecular da Célula**. 4ª Ed. Artmed SA. Porto Alegre, 2004.

ALONSO, C. R. Nonsense-mediated RNA decay: a molecular system micromanaging individual gene activities and suppressing genomic noise. **Bioessays**, v. 27, n. 5, p. 463-466, 2005.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, p. 403-410, 1990.

ANDERSSON, A.F.; LUNDGREN, M.; ERIKSSON, S.; ROSENLUND, M.; BERNANDER, R.; NILSSON, P. Global analysis of mRNA stability in the archaeon *Sulfolobus*. **Genome Biology**, v. 7, n. 10, p. R99, 2006.

ANDRADE, L. O.; ANDREWS, N. W. The *Trypanosoma cruzi* – host-cell interplay: location, invasion, retention. **Nature Reviews Microbiology**, v. 3, n. 10, p. 819-823, 2005.

ARGAMAN, M.; ALY, R.; SHAPIRA, M. Expression of heat shock protein 83 in *Leishmania* is regulated post-transcriptionally. **Molecular and Biochemical Parasitology**, v. 64, n.1, p. 95-110, 1994.

ATWOOD III, J. A.; WEATHERLY, D. B.; MINNING, T. A.; BUNDY, B.; CAVOLA, C.; OPPERDOES, F. R.; ORLANDO, R.; TARLETON, R. L. The *Trypanosoma cruzi* proteome. **Science**, v. 309, n. 5733, p. 473-476, 2005.

BACCHI, C.J.; GAROFALO, J.; SANTANA, A.; HANNAN, J.C.; BITONTI, A.J.; McCANN, P.P. *Trypanosoma brucei brucei*: regulation of ornithine decarboxylase in procyclic forms and trypomastigotes. **Experimental Parasitology**, v. 68, n.4, p. 392-402, 1989.

BARRETT, M. P.; BURCHMORE, R. J. S.; STICH, A.; LAZZARI, J. O.; FRASCH, A. C.; CAZZULO, J. J.; KRISHNA, S. The trypanosomiases. **Lancet**, v. 362, n. 1, p. 1469-1480, 2003.

BENNE, R. RNA editing in trypanosomes. **European Journal of Biochemistry / FEBS (Federation of European Biochemical Societies)**, v. 221, n. 1, p. 9-23, 1994.

BERNSTEIN, J.A.; KHODURSKY, A.B.; LIN, P.H.; LIN-CHAO, S.; COHEN, S.N. Global analysis of mRNA decay and abundance in *Escherichia coli* at a single-gene resolution using two color fluorescent DNA microarrays. **Proceedings of the National Academy of Sciences (EUA)**, v. 99, p. 9697-9702, 2002.

BONALDO, M. C.; SOUTO-PADRON, T.; DE SOUZA, W.; GOLDENBERG, S. Cell-substrate adhesion during *Trypanosoma cruzi* differentiation. **The Journal of Cell Biology**, v. 106, n. 4, p. 1349-1358, 1988.

BRISSE, S., BARNABE, C., TIBAYRENC, M. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. **International Journal for Parasitology**, v. 30, n. 1, p. 35-44, 2000.

BRITTINGHAM, A.; MILLER, M.A.; DONELSON, J.E.; WILSON, M.E. Regulation of gp63 mRNA stability in promastigotes of virulent and attenuated *Leishmania chagasi*. **Molecular and Biochemical Parasitology**, v. 112, n. 1, p. 51-59, 2001.

BUSCAGLIA, C. A.; DI NOIA, J. M. *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. **Microbes and Infection**, v. 5, n. 5, p. 419-427, 2003.

CAMARGO, E. P. Growth and differentiation in *Trypanosoma cruzi*: Origin of metacyclic trypomastigotes in liquid media. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 6, p. 93-100, 1964.

CAMPBELL, D. A.; THOMAS, S.; STURM, N. R. Transcription in kinetoplastid protozoa: why be normal? **Microbes and Infection**, v. 5, n. 13, p. 1231-1240, 2003.

CAVALIER-SMITH, T. Kingdom Protozoa and its 18 phyla. **Microbiological Reviews**, v. 57, n. 4, p. 953-994, 1993.

CAVALIER-SMITH, T. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. **Biology Letters**, v. 6, n. 3, p. 342-345, 2010.

CHAGAS, C. Nova tripanosomíase humana. Estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade mórbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, p. 159-218, 1909.

CHANG, Y. F.; IMAM, J. S.; WILKINSON, M. F. The nonsense-mediated decay RNA surveillance pathway. **Annual Review of Biochemistry**, v.76, p. 51-74, 2007.

CHAPPUIS, F.; SUNDAR, S.; HAILU, A.; GHALIB, H.; RIJAL, S.; PEELING, R. W.; ALVAR, J.; BOELAERT, M. Visceral leishmaniasis: what are the needs for diagnosis, treatment and control? **Nature Reviews Microbiology**, v. 5, n. 11, p. S7 – S16, 2007.

CHAUDHURI, M.; SHARAN, R.; HILL, G. C. Trypanosome alternative oxidase is regulated post-transcriptionally at the level of RNA stability. **The Journal of Eukaryotic Microbiology**, v. 49, n. 4, p. 263-269, 2002.

CLAYTON, C. E. Life without transcriptional control? From fly to man and back again. **The EMBO Journal**, v. 21, n. 8, p. 1881-1888, 2002.

CLAYTON, C.; SHAPIRA, M. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. **Molecular and Biochemical Parasitology**, v. 156, n. 2, p. 93-101, 2007.

COLASANTE, C.; ROBLES, A.; LI, C.H.; SCHWEDE, A.; BENZ, C.; VONCKEN, F.; GUILBRIDE, D. L.; CLAYTON, C. Regulated expression of glycosomal phosphoglycerate kinase in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 151, n. 2, p. 193-204, 2007.

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J. M.; TEROL, J.; TALÓN, M.; ROBLES, M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674-3676, 2005.

CONTRERAS, V. T.; SALLES, J. M.; THOMAS, N.; MOREL, C. M.; GOLDENBERG, S. *In vitro* differentiation of *Trypanosoma cruzi* under chemically defined conditions. **Molecular and Biochemical Parasitology**, v. 16, n. 3, p. 315-327, 1985.

CONTRERAS, V. T.; ARAUJO-JORGE, T. C.; BONALDO, M. C.; THOMAS, N.; BARBOSA, H. S.; MEIRELLES, M. N. L.; GOLDENBERG, S. Biological aspects of the Dm28c clone of *Trypanosoma cruzi* after metacyclogenesis in chemically defined media. **Memórias do Instituto Oswaldo Cruz**, v. 83, p. 123-133, 1988.

COUGHLIN, B. C.; TEIXEIRA, S. M.; KIRCHHOFF, L. V.; DONELSON, J. E. Amastin mRNA abundance in *Trypanosoma cruzi* is controlled by a 3'-untranslated region position-dependent cis-element and an untranslated region-binding protein. **The Journal of Biochemical Chemistry**, v. 275, n. 16, p. 12051-12060, 2000.

COURA J. R.; DE CASTRO S. L. A Critical Review on Chagas Disease Chemotherapy. **Memórias do Instituto Oswaldo Cruz**, v. 97, n. 1, p. 3-24, 2002.

COURA, J. R. Transmission of chagasic infection by oral route in the natural history of Chagas' disease. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 39, Suppl. 3, p. 113-117, 2006.

COURA, J. R. Chagas disease: what is known and what is needed – A background article. **Memórias do Instituto Oswaldo Cruz**, v. 102, Suppl. 1, p. 113-122, 2007.

DA CRUZ, M. Q.; BRASCHER, H. M.; VARGENS, J. R.; OLIVEIRA-LIMA, A. Effect of actinomycin D on *Trypanosoma cruzi*. **Experientia**, v. 47, n. 1, p. 89-92, 1991.

DALLAGIOVANNA, B.; CORREA, A.; PROBST, C. M.; HOLETZ, F.; SMIRCICH, P.; DE AGUIAR, A. M.; MANSUR, F.; DA SILVA, C. V.; MORTARA, R. A.; GARAT, B.; BUCK, G. A.; GOLDENBERG, S.; KRIEGER, M. A. Functional genomic characterization of mRNAs associated with TcPUF6, a pumilio-like protein from *Trypanosoma cruzi*. **The Journal of Biological Chemistry**, v. 283, n. 13, p. 8266-8273, 2008.

DECKER, C. J. The exosome: a versatile RNA processing machine. **Current Biology**, v. 8, n. 7, p. R238-R240, 1998.

DeRISI, J. L.; VISHWANATH, R. I.; BROWN, P. O. Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale. **Science**, v. 278, p. 680-686, 1997.

DE SOUZA, W. O Parasito e sua Interação com os Hospedeiros. In: BRENNER, Z.; ANDRADE, A. Z.; BARRAL-NETTO, M. (eds). **Trypanosoma cruzi e Doença de Chagas**. 2.^a Ed. Guanabara Koogan SA. Rio de Janeiro, 2000.

DEVERA, R.; FERNANDES, O.; COURA, J. R. Should *Trypanosoma cruzi* be called "cruzi" complex? A review of the parasite diversity and the potential of selecting population after *in vitro* culturing and mice infection. **Memórias do Instituto Oswaldo Cruz**, v. 98, n. 1, p. 1-12, 2003.

DIAS, J. C. P. Epidemiologia. *In*: BRENNER, Z.; ANDRADE, A. Z.; BARRAL-NETTO, M. (eds). ***Trypanosoma cruzi e Doença de Chagas***. 2.^a Ed. Guanabara Koogan SA. Rio de Janeiro, 2000.

DJIKENG, A.; FERREIRA, L.; D'ANGELO, M.; DOLEZAL, P.; LAMB, T.; MURTA, S.; TRIGGS, V.; ULBERT, S.; VILLARINO, A.; RENZI, S.; ULLU, E; TSCHUDI, C. Characterization of a candidate *Trypanosoma brucei* U1 small nuclear RNA gene. **Molecular and Biochemical Parasitology**, v. 133, n. 1, p. 109-115, 2001.

DOSSIN, F. de M.; SCHENKMAN, S. Actively transcribing RNA polymerase II concentrates on spliced leader genes in the nucleus of *Trypanosoma cruzi*. **Eukaryotic Cell**, v. 4, n. 5, p. 960-970, 2005.

DUDOIT, S.; GENTLEMAN, R. C.; QUACKENBUSH, J. Open source software for the analysis of microarray data. **Biotechniques**, Suppl., p. 45-51, 2003.

DUMONTEIL, E. DNA Vaccines against Protozoan Parasites: Advances and Challenges. **Journal of Biomedicine and Biotechnology**, v. 2007, 2007:90520, 2007.

EHLERS, B.; CZICHOS, J.; OVERATH, P. RNA turnover in *Trypanosoma brucei*. **Molecular and Cellular Biology**, v. 7, n. 3, p. 1242-1249, 1987.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences (EUA)**, v. 95, n. 25, p. 14863-14868, 1998.

ELIAS, M. C.; MARQUES-PORTO, R.; FREYMÜLLER, E.; SCHENKMAN, S. Transcription rate modulation through the *Trypanosoma cruzi* life cycle occurs in parallel with changes in nuclear organisation. **Molecular and Biochemical Parasitology**, v. 112, n. 1, p. 79-90, 2001.

EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C.; NILSSON, D.; AGGARWAL, G.; TRAN, A. N.; GHEDIN, E.; WORTHEY, E. A.; DELCHER, A. L.; BLANDIN, G.; WESTENBERGER, S. J. *et al.* The genome sequence of

Trypanosoma cruzi, etiologic agent of Chagas disease. **Science**, v. 309, n. 5733, p. 409-415, 2005(a).

EL-SAYED, N. M.; MYLER, P. J.; BLANDIN, G.; BERRIMAN, M.; CRABTREE, J.; AGGARWAL, G.; CALER, E.; RENAULD, H.; WORTHEY, E. A.; HERTZ-FOWLER, C.; GHEDIN, E.; PEACOCK, C. *et al.* Comparative genomics of trypanosomatid parasitic protozoa. **Science**, v. 309, n. 5733, p. 404-409, 2005(b).

ESTÉVEZ, A. M.; KEMPF, T.; CLAYTON, C. The exosome of *Trypanosoma brucei*. **The EMBO Journal**, v. 20, n. 14, p. 3831-3839, 2001.

FERREIRA, L. R.; DOSSIN, F. DE M.; RAMOS, T. C.; FREYMÜLLER, E.; SCHENKMAN, S. Active transcription and ultrastructural changes during *Trypanosoma cruzi* metacyclogenesis. **Anais da Academia Brasileira de Ciências**, v. 80, n. 1, p. 157-66, 2008.

GARCIA, A.; COURTIN, D.; SOLANO, P.; KOFFI, M.; JAMONNEAU, V. Human African trypanosomiasis: connecting parasite and host genetics. **Trends in Parasitology**, v. 22, n. 9, p. 405-409, 2006.

GARG, N.; POPOV, V. L.; PAPACONSTANTINO, J. Profiling gene transcription reveals a deficiency of mitochondrial oxidative phosphorylation in *Trypanosoma cruzi*-infected murine hearts: implications in chagasic myocarditis development. **Biochimica et Biophysica Acta**, v. 1638, n. 2, p. 106-120, 2003.

GENTIL, L. G.; CORDERO, E. M.; DO CARMO, M. S.; DOS SANTOS, M. R.; DA SILVEIRA, J. F. Posttranscriptional mechanisms involved in the control of expression of the stage-specific GP82 surface glycoprotein in *Trypanosoma cruzi*. **Acta Tropica**, v. 109, n. 2, p. 152-158, 2009.

GENTLEMAN, R. C.; CAREY, V. J.; BATES, D. M.; BOLSTAD, B.; DETTLING, M.; DUDOIT, S.; ELLIS, B.; GAUTIER, L.; GE, Y.; GENTRY, J.; HORNIK, K.; HOTHORN, T.; HUBER, W.; IACUS, S.; IRIZARRY, R.; LEISCH, F.; LI, C.; MAECHLER, M.; ROSSINI, A. J.; SAWITZKI, G.; SMITH, C.; SMYTH, G.; TIERNEY, L.; YANG, J. Y.; ZHANG, J. Bioconductor: open software development for computational biology and bioinformatics. **Genome Biology**, v. 5, n. 10, p.R80, 2004.

GILLINGWATER, K.; BÜSCHER, P.; BRUN, R. Establishment of a panel of reference *Trypanosoma evansi* and *Trypanosoma equiperdum* strains for drug screening. **Veterinary Parasitology**, v. 148, n. 2, p. 114-121, 2007.

GLISOVIC, T.; BACHORIK, J. L.; YONG, J.; DREYFUSS, G. RNA-binding proteins and post-transcriptional gene regulation. **FEBS (Federation of European Biochemical Societies) Letters**, v. 582, n. 14, p. 1977-1986, 2008.

GODOI, M. M. I. M. ***Phytomonas* e outros Tripanossomatídeos em insetos no Estado de Rondônia** – Amazônia Ocidental. São Paulo, 2000. Dissertação (Mestrado em Ciências) – Instituto de Ciências Biomédicas da Universidade de São Paulo.

GRYNBERG, P.; PASSOS-SILVA, D. G.; MOURÃO, M. DE M.; HIRATA, R. JR.; MACEDO, A. M.; MACHADO, C. R.; BARTHOLOMEU, D. C.; FRANCO, G. R. *Trypanosoma cruzi* Gene Expression in Response to Gamma Radiation. **PLoS One**, v. 7, n. 1, p. e29596, 2012.

HAMBRAEUS, G. VON WACHENFELDT, C. HEDERSTEDT, L. Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. **Molecular Genetics and Genomics**, v. 269, n. 5, p. 706-14, 2003.

HARTLEY, J. L.; TEMPLE, G. F.; BRASCH, M. A. DNA cloning using *in vitro* site-specific recombination. **Genome Research**, v. 10, n. 11, p. 1788-1795, 2000.

HENTZE, M. W.; KULOZIK, A. E. A perfect message: RNA surveillance and nonsense-mediated decay. **Cell**, v. 96, n. 3, p. 307-310, 1999.

HONIGBERG, B. A contribution to systematics of the nonpigmented flagellates. *In* **Progress in Protozoology** (Ludvík, J. *et al.*, eds), p. 68–69, Czechoslovak Academy of Sciences, 1963.

HOU, M. H.; ROBINSON, H.; GAO, Y. G.; WANG, A. H. Crystal structure of actinomycin D bound to the CTG triplet repeat sequences linked to neurological diseases. **Nucleic Acids Research**, v. 30, n. 22, p. 4910-4917, 2002.

HOUSELEY, J.; LACAVA, J.; TOLLERVEY, D. RNA-quality control by the exosome. **Nature Reviews Molecular Cell Biology**, v. 7, n. 7, p. 529-539, 2006.

HUYNEN, M. A.; BORK, P. Measuring genome evolution. **Proceedings of the National Academy of Sciences (EUA)**, v. 95, n. 11, p. 5849-5856, 1998.

JONES, T. W.; DÁVILA, A. M. R. *Trypanosoma vivax* – out of África. **Trends in Parasitology**, v. 17, n. 2, p. 99-101, 2001.

JUNQUEIRA, A. C.; DEGRAVE, W.; BRANDÃO, A. Minicircle organization and diversity in *Trypanosoma cruzi* populations. **Trends in Parasitology**, v. 21, n. 6, p. 270-272, 2005.

KADO, C. I. Horizontal gene transfer: sustaining pathogenicity and optimizing host-pathogen interactions. **Molecular Plant Pathology**, v. 10, n. 1, p. 143-150, 2009.

KEENE, J. D. RNA regulons: coordination of post-transcriptional events. **Nature Reviews Genetics**, v. 8, n. 7, p. 533-543, 2007.

KESSLER, R. L. **Transcriptômica de *Trypanosoma cruzi* em resposta a inibidores da síntese de esteróis**. Curitiba, 2010. Dissertação (Mestrado em Biologia Celular e Molecular) – Universidade Federal do Paraná.

KNOLL, A. H.; CARROLL, S. B. Early animal evolution: emerging views from comparative biology and geology. **Science**, v. 284, n. 5423, p. 2129-2137, 1999.

KRIEGER, M. A.; SALLES, J. M.; ALMEIDA, E.; LINSS, J.; BONALDO, M. C.; GOLDENBERG, S. Expression and polymorphism of a *Trypanosoma cruzi* gene encoding a cytoplasmic repetitive antigen. **Molecular and Biochemical Parasitology**, v. 70, n. 3, p. 247-254, 1990.

KOBA, M.; KONOPA, J. Actinomycin D and its mechanisms of action. **Postępy higieny i medycyny doświadczalnej** (O progresso da higiene e da medicina experimental), v. 59, p. 290-298, 2005.

LANDER, E. S. Array of hope. **Nature Genetics**, v.21, n. 1, p. 3-4, 1999.

LANDY, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. **Annual Review of Biochemistry**, v. 58, p. 913-949, 1989.

MACEDO, A. M.; OLIVEIRA, R. P.; PENA, S. D. Chagas disease: role of parasite genetic variation in pathogenesis. **Expert Reviews in Molecular Medicine**, v. 4, n. 5, p. 1-16, 2002.

MAIR, G.; SHI, H.; LI, H.; DJIKENG, A.; AVILES, H.O.; BISHOP, J. R.; FALCONE, F. H.; GAVRILESCU, C.; MONTGOMERY, J. L.; SANTORI, M. I.; STERN, L. S.; WANG, Z.; ULLU, E.; TSCHUDI, C. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. **RNA**, v. 6, n. 2, p. 163-169, 2000.

MARIS, C.; DOMINGUEZ, C.; ALLAIN, F. H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. **The FEBS (Federation of European Biochemical Societies) Journal**, v. 272, n. 9, p. 2118-2131, 2005.

MAYR, E. W. Emergência. *In: O Desenvolvimento do Pensamento Biológico: diversidade, evolução e herança*. 1ª ed. Brasília. Editora da Universidade de Brasília, 1998. p. 83-87.

MINNING, T. A.; WEATHERLY, D. B.; ATWOOD III, J. A.; ORLANDO, R.; TARLETON, R. L. The steady-state transcriptome of the four major life-cycle stages of *Trypanosoma cruzi*. **BMC Genomics**, v. 10, p. 370, 2009.

MINNING, T. A.; WEATHERLY, D. B.; FLIBOTTE, S.; TARLETON, R. L. Widespread, focal copy number variations (CNV) and whole chromosome aneuploidies in *Trypanosoma cruzi* strains revealed by array comparative genomic hybridization. **BMC Genomics**, v. 12, p. 139, 2011.

MITCHELL, P.; PETFALSKI, E.; SHEVCHENKO, A.; MANN, M.; TOLLERVEY, D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. **Cell**, v. 91, n. 4, p. 457-466, 1997.

MUKHOPADHYAY, R.; MADHUBALA, R. Leishmania donovani: cellular control of ornithine decarboxylase in promastigotes. **International Journal of Biochemistry and Cell Biology**, v. 27, n. 9, p. 947-952, 1995.

NÁZER, E.; VERDÚN, R. E.; SÁNCHEZ, D. O. Nucleolar localization of RNA binding proteins induced by actinomycin D and heat shock in *Trypanosoma cruzi*. **PLoS One**, v. 6, n. 5, p. e19920, 2011.

NÁZER, E.; SÁNCHEZ, D. O. Nucleolar accumulation of RNA binding proteins induced by Actinomycin D is functional in *Trypanosoma cruzi* and *Leishmania mexicana* but not in *T. brucei*. **PLoS One**, v. 6, n. 8, p. e24184, 2011.

NEVES, D. P. **Parasitologia Humana**. 8ª Ed. Atheneu. São Paulo, 1991.

PALFI, Z.; LANE, W. S.; BINDEREIF, A. Biochemical and functional characterization of the cis-spliceosomal U1 small nuclear RNP from *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 121, n. 2, p. 233-243, 2002.

PATTERSON, D. J.; LARSEN, J. (eds). **The Biology of Free-Living Heterotrophic Flagellates**. Editora Clarendon Press. Oxford, 1991.

PAVONI, D. P. **Caracterização genômica funcional das formas infectivas do *Trypanosoma cruzi***. Curitiba, 2005. Tese (Doutorado em Biologia Celular e Molecular) - Instituto Oswaldo Cruz.

PAYS, E.; HANOCQ-QUERTIER, J.; HANOCQ, F.; VAN ASSEL, S.; NOLAN, D.; ROLIN, S. Abrupt RNA changes precede the first cell division during the differentiation of *Trypanosoma brucei* bloodstream forms into procyclic forms *in vitro*. **Molecular and Biochemical Parasitology**, v. 61, n. 1, p.107-114, 1993.

PROBST, C. M. **Descrição da metaciclologênese de *Trypanosoma cruzi* pelo uso de microarranjos de DNA**. Curitiba, 2005. Tese (Doutorado em Biologia Celular e Molecular) - Instituto Oswaldo Cruz.

PUIG, O.; CASPARY, F.; RIGAUT, G.; RUTZ, B.; BOUVERET, E.; BRAGADO-NILSSON, E.; WILM, M.; SÉRAPHIN, B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. **Methods**, v. 24, n. 3, p. 218-229, 2001.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org/>, 2010.

RAGHAVAN, A.; OGILVIE, R.L.; REILLY, C.; ABELSON, M.L.; RAGHAVAN, S.; VASDEWANI, J.; KRATHWOHL, M.; BOHJANEN, P.R. Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. **Nucleic Acids Research**, v. 30, n. 2, p. 5529-5538, 2002.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.

ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biology**, v. 11, n. 3, p. R25, 2010.

SCHENA, M.; SHALON, D.; DAVIS, R. W.; BROWN, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, v. 270, n. 5235, p. 467-470, 1995.

SCHULZE, A.; DOWNWARD, J. Navigating gene expression using microarrays – a technology review. **Nature Cell Biology**, v. 3, n. 8, p. E190-E195, 2001.

SHENDURE, J.; JI, H. Next-generation DNA sequencing. **Nature Biotechnology**, vol. 26, n. 10, p. 1135-1145, 2008.

SHOCK, J. L.; FISCHER, K. F.; DeRISI, J. L. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. **Genome Biology**, v. 8, n. 7, R134, 2007.

SIEGEL, T. N.; HEKSTRA, D. R.; WANG, X.; DEWELL, S.; CROSS, G. A. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. **Nucleic Acids Research**, v. 38, n. 15, p. 4946-4957, 2010.

SIMPSON, A. G. B.; STEVENS, J. R.; LUKEŠ, J. The evolution and diversity of kinetoplastid flagellates. **Trends in Parasitology**, v. 22, n. 4, p. 168-174, 2006.

SOBELL, H. M. Actinomycin and DNA transcription. **Proceedings of the National Academy of Sciences (EUA)**, v. 82, n.16, p.5328-5331, 1985.

SOKAL, R. R.; MICHENER, C. D. A Statistical Method for Evaluating Systematic Relationships. **The University of Kansas Scientific Bulletin**, v. 38, p. 1409-1438, 1958.

SMITH, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. **Statistical applications in Genetics and Molecular Biology**, v. 3, n. 1, art. 3, 2004.

SPELLMAN, P. T.; SHERLOCK, G.; ZHANG, M. Q.; IYER, V. R.; ANDERS, K.; EISEN, M. B.; BROWN, P. O.; BOTSTEIN, D.; FUTCHER, B. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. **Molecular Biology of the Cell**, v. 9, n. 12, p. 3273-3297, 1998.

SUNNERHAGEN, P. Cytoplasmatic post-transcriptional regulation and intracellular signalling. **Molecular Genetics and Genomics**, v. 277, n. 4, p. 341-355, 2007.

TDR (Tropical Disease Research Programme) Disease Watch. **Nature Reviews Microbiology**, v. 1, n. 1, p. 14-15, 2003.

TEIXEIRA, A. R.; NASCIMENTO, R. J.; STURM, N. R. Evolution and pathology in Chagas disease – A Review. **Memórias do Instituto Oswaldo Cruz**, v. 101, n. 5, p. 463-491, 2006.

TEIXEIRA, M. M. G.; BORGHESAN, T. C.; FERREIRA, R. C.; SANTOS, M. A.; TAKATA, C. S. A.; CAMPANER, M.; NUNES, V. L. B.; MILDER, R. V.; de SOUZA, W.; CAMARGO, E. P. Phylogenetic validation of the Genera *Angomonas* and *Strigomonas* of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. **Protist**, v. 162, n. 3, p. 503-524, 2011.

THOMAS, M. C.; GARCÍA-PÉREZ, J. L.; ALONSO, C.; LÓPEZ, M. C. Molecular characterization of KMP-11 from *Trypanosoma cruzi*: a cytoskeleton-associated protein regulated at the translational level. **DNA and Cell Biology**, v. 19, n. 1, p. 47-57, 2000.

TIBAYRENC, M. Genetic subdivisions within *Trypanosoma cruzi* (Discrete Typing Units) and their relevance for molecular epidemiology and experimental evaluation. **Kinetoplastid Biology and Disease**, v. 2, n. 1, p. 12, 2003.

TRAVIS, J. A close look at urbisexuality. **Science**, v. 316, n. 5823, p. 390-391, 2007.

TYLER, K. M.; ENGMAN, D. M. The life cycle of *Trypanosoma cruzi* revisited. **International Journal for Parasitology**, v. 31, n. 5-6, p. 472-481, 2001.

VAENA DE AVALOS, S.; BLADER, I. J.; FISHER, M.; BOOTHROYD, J. C.; BURLEIGH, B. A. Immediate/early response to *Trypanosoma cruzi* infection involves minimal modulation of host cell transcription. **The Journal of Biological Chemistry**, v. 277, n. 1, p. 639-644, 2002.

VANHAMME, L.; PAYS, E. Control of gene expression in trypanosomes. **Microbiological Reviews**, v. 59, n. 2, p. 223-240, 1995.

VANHAMME, L.; POSTIAUX, S.; POELVOORDE, P.; PAYS, E. Differential regulation of ESAG transcripts in *Trypanosoma brucei*. **Molecular and Biochemical Parasitology**, v. 102, n. 1, p. 35-42, 1999.

VELCULESCU, V. E.; ZHANG, L. ; VOGELSTEIN, B.; KINZLER, K. W. Serial analysis of gene expression, **Science**, vol. 270, n. 5235, p. 484-487, 1995.

VIDAL, M. A biological atlas of functional maps. **Cell**, v. 104, n. 3, p. 333-339, 2001.

VIDAL, M. A unifying view of 21st century systems biology. **FEBS (Federation of European Biochemical Societies) Letters**, v. 583, n. 24, p. 3891-3894, 2009.

VILLELA, M. M.; DE SOUZA, J. M.; MELO V. de P.; DIAS, J. C. Chagas disease epidemiological surveillance in a decentralized program: evaluation of practice and knowledge among municipal health agents in an endemic region of Minas Gerais State, Brazil. **Cadernos de Saúde Pública**, v. 23, n. 10, p. 2428-2438, 2007.

WANG, Y.; LIU, C. L.; STOREY, J. D.; TIBSHIRANI, R. J.; HERSCHLAG, D.; BROWN, P. O. Precision and functional specificity in mRNA decay. **Proceedings of the National Academy of Sciences (EUA)**, v. 99, p. 5860-5865, 2002.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews**, vol. 10, p. 57-63, 2009.

WESTON, D.; LA FLAMME, A. C.; VAN VOORHIS, W. C. Expression of *Trypanosoma cruzi* surface antigen FL-160 is controlled by elements in the 3' untranslated, the 3' intergenic, and the coding regions. **Molecular and Biochemical Parasitology**, v. 102, n. 1, p. 53-66, 1999.

YANG, E.; VAN NIMWEGEN, E.; ZAVOLAN, M.; RAJEWSKY, N.; SCHOROEDER, M.; MAGNASCO, M.; DARNELL, JE Jr. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. **Genome Research**, v. 13, p. 1863-1872, 2003.

ZAVERUCHA DO VALLE, T.; CALABRESE, K. S.; CORTE-REAL, S.; BAETAS, W. C.; GONÇALVES DA COSTA, S. C. *Trypanosoma cruzi*: *in vitro* morphological alterations induced by actinomycin D. **Pharmacology**, v. 67, n. 2, p. 55-58, 2003.

ZINGALES, B.; STOLF, B. S.; SOUTO, R. P.; FERNANDES, O.; BRIONES, M. R. S. Epidemiology, Biochemistry and Evolution of *Trypanosoma cruzi* Lineages Based on Ribosomal RNA Sequences. **Memórias do Instituto Oswaldo Cruz**, v. 94, Supl. I, p. 159-164, 1999.

ZINGALES, B.; ANDRADE, S. G.; BRIONES, M. R.; CAMPBELL, D. A.; CHIARI, E.; FERNANDES, O.; GUHL, F.; LAGES-SILVA, E.; MACEDO, A. M.; MACHADO, C. R.; MILES, M. A.; ROMANHA, A. J.; STURM, N. R.; TIBAYRENC, M.; SCHIJMAN, A. G. *et al.* New consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051-1054, 2009.